Assessing the value of 18FDG-PET in lung cancer

van Tinteren, H.

2006

document version
Publisher's PDF, also known as Version of record

ASSESSING THE VALUE OF ^{18}FDG-PET IN LUNG CANCER:

FROM THEORY TO PRACTICE

The publication of this thesis was financially supported by:
The Comprehensive Cancer Center Amsterdam (IKA),
BV Cyclotron VU,
Siemens Nederland N.V.

Assessing the value of ^{18}FDG-PET in lung Cancer: From theory to practice.

ISBN-10: 90-74946-12-7
ISBN-13: 978-90-74946-12-4

Cover design: Renato Valdés Olmos, www.atacamadesign.com
Printed by: Buijten & Schipperheijn, Amsterdam, The Netherlands

VRIJE UNIVERSITEIT

ASSESSING THE VALUE OF ^{18}FDG-PET IN LUNG CANCER: FROM THEORY TO PRACTICE

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. L.M. Bouter,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Geneeskunde
op woensdag 13 december 2006 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Harm van Tinteren

geboren te Aruba

promotoren:	prof. dr. M. Boers
	prof. dr. G. J. J. Teule
copromotor:	prof. dr. O. S. Hoekstra

Contents

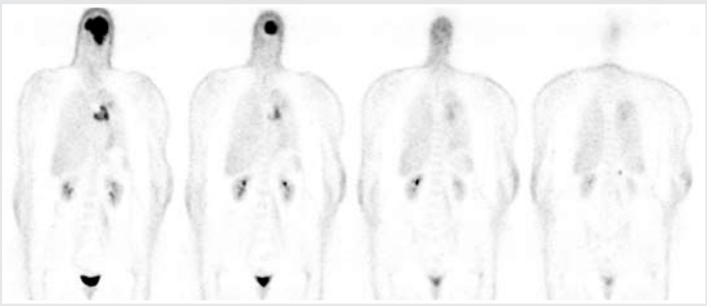
CHAPTER 1

Introduction

Positron Emission Tomography in Oncology

Positron Emission Tomography (PET) is the study and visualization of human physiology by electronic detection of short-lived positron emitting radiopharmaceuticals. It is the only non-invasive technology that can routinely and quantitatively measure metabolic, biochemical and functional activity in living tissue. This differs from other forms of imaging such as magnetic resonance imaging (MRI), computed tomography (CT) or ultrasound techniques, which mainly show structural (anatomical) information.

In cancer, the most common radiolabelled tracer is [^{18}F]-fluoro-2-deoxy-D-glucose (FDG). Cancer cells have a much higher metabolic rate than other cells. One characteristic is that cancer cells need higher levels of glucose for energy. This is the biological process PET measures. Therefore FDG-PET allows differentiation between malignant and benign abnormalities.
PET can also help physicians monitor the treatment of disease. For example, chemotherapy leads to changes in cellular activity, and that may be observable by PET long before structural changes can be measured by ultrasound, X-rays, CT, or MRI.



Legend to the figure: Typical whole-body PET image of a patient with NSCLC. Apart from uptake in the lung and the mediastinum, there is physiological uptake in the brain and bladder. The most important finding was a metastases near the left kidney (most right image). Due to this finding, the planned policy (neoadjuvant chemotherapy followed by surgery) changed into palliative management.

Introduction

In 2003 the results were published of a randomized controlled trial (RCT) comparing magnetic resonance imaging (MRI) with plain radiographs in patients that were referred to the hospital for radiographic evaluation of low back pain.[1] Improvements in the MRI-technique had made the comparison a clinically relevant issue. Both imaging techniques resulted in nearly identical outcomes of disability, pain and general health status for patients. However, MRI increased the costs of care because of the higher costs of the technique itself and the increased number of operations and specialist consultations due to MRI. Although physicians preferred using MRI, substituting MRI for radiographs in primary care of low back pain is not indicated.

PET Scans a Clinical Reality

If you could manage and treat your cancer patients more accurately 45% of the time[1], you might call it a clinical breakthrough…
We call it PET.

[1] Tucker et al., *Journal of Clinical Oncology* – Vol. 19, No. 9, 2001, pp. 2504-2508.

FIGURE 1. Source: http://www.petscaninfo.com

In the '90s, ^{18}F-fluorodeoxyglucose (FDG)-positron emission tomography (PET) emerged as a promising innovative imaging technique (see inlay). However, the technique is expensive and for the time being its availability is limited. Therefore, to become of benefit for patients and society, PET needs to be evaluated together and in contrast to other non-invasive techniques to define its most efficient indications.

In general, the diagnosis of a disease, for example of a particular cancer, involves a multidisciplinary, step-by-step process. The goal of that process is to determine the most likely status of disease in order to give the patient the therapy that is most appropriate to that disease condition. Each test contributes a piece of information. In diagnostic accuracy studies the information is usually expressed in a two-by-two table. The test is either positive or negative and the disease is present or absent. A series of statistics are based on these combinations. Sensitivity, the number of true test positives divided by all cases with disease and specificity, the number of true test negatives divided by all cases without disease, are the most common accuracy measures. In reality however, unless it is based on a pathognomonic sign of disease, a test will rarely result in such black-and-white condition. Recently, a radiologist, forced to express himself in a language that was not his, phrased it as " Sitting on the fence – a radiologist's stock in trade – necessitates using words for balance, weighing diagnostic probabilities, and leaning toward the heavier side. But because I couldn't use the subjunctive mood, I was forced into the realm of apparent diagnostic certainty".[2] The result of a FDG-PET-scan is usually expressed qualitatively, e.g. in terms of normal, faint, moderate and intense. Translated into a diagnosis these qualifications should be interpreted somewhere between definitively benign, probably benign, equivocal, probably malignant or definitively malignant. Obviously, the dichotomy is lost. Likewise, each individual (non-invasive) test adds some probability measure and usually only invasively obtained tissue material provides histological proof of disease – in most cases.

The following example may illustrate some of the uncertainties in a common diagnostic process: lung cancer often presents as a solitary pulmonary lesion (SPN). One third of lung nodules in patients more than 35 years old are found to be malignant. Over 50% of the radiographically indeterminable nodules resected at thoracoscopy are benign.[3] The pretest probability of cancer determines the most cost-effective strategy for diagnosis of SPN. If the probability is very low, radiographic follow-up is preferred but more advanced and invasive techniques are needed when the pretest probability increases.[4] The pretest probability may be expressed as function of patient characteristics such as age and smoking history and some radiographical parameters.[5] Typically a technique such as FDG-PET may contribute to the diagnosis of SPN because it may help to distinguish benign from malignant abnormalities. In 2001, a meta-analysis was published on 40 diagnostic accuracy studies of FDG-PET in SPN. The median sensitivity and specificity for pulmonary nodules were approximately 97% and 78% respectively.[6] That makes it a very accurate test. However, whether FDG-PET is helpful in clinical decision making and actually will change subsequent patient management cannot be ascertained with these values only, because the probability of malignancy after the test depends on its pretest probability (figure 2) and the overall interpretive analysis of the physicians, c.q. what are the acceptable consequences in terms of management.

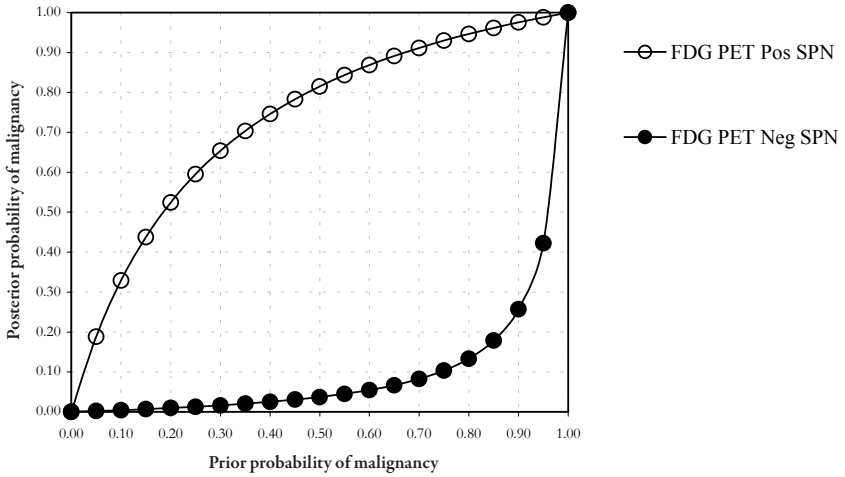


Figure 2. Posterior probability of malignancy related to prior probability using FDG-PET with a sensitivity of 97% and a specificity of 78%. Source: Gould, JAMA 2001 (6).

Staging of non-small-cell lung cancer and FDG-PET

The use of FDG-PET encompasses more potential benefits in a patient undergoing non-invasive evaluation of lung cancer than can be captured in a diagnostic accuracy study of solitary pulmonary lesions. There are 2 main categories of lung cancer: small cell lung cancer and non-small cell lung cancer (NSCLC). Yearly 9000 cases of lung cancer are diagnosed in the Netherlands, and of these 80% have non-small cell lung cancer.[7] Accurate staging of NSCLC cancer is very important because treatment options and the prognosis differ significantly by stage. Current diagnostic strategies are primarily aimed at selecting patients who are likely to benefit from surgery or multimodality treatment. A recent study based on two hospitals in our own region and a large study from The USA indicated that the pre-operative diagnostic process is suboptimal and too many patients are given the benefit of the doubt with surgery.[8,9] Already at its introduction, FDG-PET appeared promising as a technique to improve the overall diagnostic process especially with respect to reducing unnecessary surgery. It proved more accurate than CT in detecting solitary pulmonary lesions,[4,6] in detecting or evaluating mediastinal lymph nodes[10,11] and in identifying distant metastasis. However, improved accuracy in different indications does not necessarily imply overall clinical usefulness, e.g. better patient management and improved clinical outcome. To prove (cost-)-effectiveness, a series of conditions have to be met. First, the current diagnostic strategy should have shortcomings and the new test should likely be able to contribute relevant information to those deficiencies. Second, the information from the new test should be relevant and sufficiently convincing to direct the clinician to alternative patient management. Such a change in management should in turn lead to an improved patient outcome at reasonable costs.

The first condition, the potential residual inefficiency in daily clinical practice of preoperative staging in patients with suspected NSCLC was studied in two hospitals, one general and one academic institute, in our region of interest. It was found that nearly 50% of operated patients, surgical intervention failed because of irresectable tumor or benign lesion at surgery or because the disease recurred or metastasized within one year.[8] The second condition, the amount of relevant information added by the new test would be best answered by a comparison of the concurrent strategy with a strategy that includes the new test, e.g. FDG-PET as add-on. To minimize potential bias in that comparison, due to measurable and non-measurable patient- and tumor characteristics and diagnostician/clinician decisions, a randomized controlled trial is preferred, whenever feasible and ethical. Concurrent collection of information on costs in combination with decisions analysis can be used to prove cost-effectiveness of the new technology. Ultimately, a comparison of the situation before and after the introduction of a new technology after its implementation in routine care will allow a full assessment of the benefits and costs for patients and society.

In other words, the assessment of the contribution of a diagnostic test to the patient management process is a multi-phase hierarchical process, as has been described by Fryback and others.[12] The process starts with the technical aspects of the test such as image quality and reproducibility (level 1) and is followed by the assessment of the diagnostic accuracy (level 2). Next, at the third level it is investigated whether the information produces indeed changes in the physicians' diagnostic thinking. Such a change is a prerequisite for a change in patient management (level 4). At the fifth level, the actual size of effect on patient outcome is measured. And finally, the cost-benefit analysis of the introduction of the technology is made (level 5). Clearly, demonstration of efficacy at each lower level is logically necessary, but not sufficient, to assure efficacy at a higher level.

Staging NSCLC

The system most often used to describe the growth and spread of non-small cell lung cancer (NSCLC) is the **TNM** staging system, also known as the American Joint Committee on Cancer (AJCC) system. In TNM staging, information about the tumor, nearby lymph nodes, and distant organ metastases is combined and a stage is assigned to specific TNM groupings. The grouped stages are described using the number 0 and Roman numerals from I to IV.
T stands for tumor (its size and how far it has spread within the lung and to nearby organs), **N** stands for spread to lymph nodes, and **M** is for metastasis (spread to distant organs).
Stage grouping for non-small cell lung cancer: Once the T, N, and M categories have been assigned, this information is combined (stage grouping) to assign an overall stage of 0, I, II, III, or IV. Patients with lower stage numbers have a more favorable outlook for survival.

Overall Stage	T category	N category	M category
Stage 0	Tis (In situ)	N0	M0
Stage IA	T1	N0	M0
Stage IB	T2	N0	M0
Stage IIA	T1	N1	M0
Stage IIB	T2	N1	M0
	T3	N0	M0
Stage IIIA	T1	N2	M0
	T2	N2	M0
	T3	N1	M0
	T3	N2	M0
Stage IIIB	Any T	N3	M0
	T4	Any N	M0
Stage IV	Any T	Any N	M1

The thesis

This thesis is about the design of studies to evaluate PET as a novel diagnostic technology with emphasis on its potential role in NSCLC. The hierarchical approach act as the theoretical backbone and improved diagnostic accuracy of PET in NSCLC has been considered as point of departure. To explore and improve our understanding of the potential impact of FDG-PET on management decisions in the pre-operative setting, we first performed a 'clinical value' or before-after' study in patients referred to the PET-center with a diagnostic NSCLC problem. The results are presented in *chapter 2*. For the purpose of designing the diagnostic RCT, we performed a literature search (in 1996, which was reiterated in 2005) to find other randomized diagnostic imaging studies. Details of this search are described in *chapter 3*. Meanwhile, we designed our first RCT and the essentials of the protocol and the logistics are explained in *chapter 4*. The study got the name 'PET in LUng cancer Staging' (PLUS) and in *chapter 5* the results of the PLUS study are shown. Because the data on costs of the work-up were collected concurrently, we were able to calculate the overall costs. In addition, different scenarios depending on tracer and scanning capacity were explored by means of sensitivity analysis in *chapter 6*. The need for randomization in a comparison of different diagnostic strategies to determine the added value of a diagnostic device is not appreciated by every researcher. In *chapter 7* we formulated our point of view with respect to the role of randomization in diagnostic research. Finally, in *chapter 8* an overview is given of studies undertaken along the hierarchy of the theoretical framework to evaluate FDG-PET. A summary of the thesis and some discussion on current developments can be found in *chapter 9*.

Reference List

1. Jarvik JG, Hollingworth W, Martin B et al. Rapid magnetic resonance imaging vs radiographs for patients with low back pain: a randomized controlled trial. *JAMA* 2003; 289(21):2810-2818.
2. Bruzzi JF. The words count – radiology and medical linguistics. *N Engl J Med* 2006; 354(7):665-667.
3. Mack MJ, Hazelrigg SR, Landreneau RJ, Acuff TE. Thoracoscopy for the diagnosis of the indeterminate solitary pulmonary nodule. *Ann Thorac Surg* 1993; 56(4):825-830.
4. Gambhir SS, Shepherd JE, Shah BD et al. Analytical decision model for the cost-effective management of solitary pulmonary nodules. *J Clin Oncol* 1998; 16(6):2113-2125.
5. Swensen SJ, Silverstein MD, Ilstrup DM, Schleck CD, Edell ES. The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules. *Arch Intern Med* 1997; 157(8):849-855.
6. Gould MK, Maclean CC, Kuschner WG, Rydzak CE, Owens DK. Accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions: a meta-analysis. *JAMA* 2001; 285(7):914-924.
7. Siesling S, van Dijck JA, Visser O, Coebergh JW. Trends in incidence of and mortality from cancer in The Netherlands in the period 1989-1998. *Eur J Cancer* 2003; 39(17):2521-2530.
8. Herder GJ, Verboom P, Smit EF et al. Practice, efficacy and cost of staging suspected non-small cell lung cancer: a retrospective study in two Dutch hospitals. *Thorax* 2002; 57(1):11-14.
9. Little AG, Rusch VW, Bonner JA et al. Patterns of surgical care of lung cancer patients. *Ann Thorac Surg* 2005; 80(6):2051-2056.
10. Gould MK, Kuschner WG, Rydzak CE et al. Test performance of positron emission tomography and computed tomography for mediastinal staging in patients with non-small-cell lung cancer: a meta-analysis. *Ann Intern Med* 2003; 139(11):879-892.
11. Dwamena BA, Sonnad SS, Angobaldo JO, Wahl RL. Metastases from non-small cell lung cancer: mediastinal staging in the 1990s – meta-analytic comparison of PET and CT. *Radiology* 1999; 213(2):530-536.
12. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991; 11(2):88-94.

CHAPTER 21

Prospective use of serial questionnaires to evaluate the therapeutic efficacy of ^{18}FDG PET in (suspected) lung cancer

Gerarda J Herder[1,2], Harm van Tinteren[3], Emile F Comans[1], Otto S Hoekstra[1,4], Gerrit J Teule[1], Pieter E Postmus[2], Urvi Joshi[1], Egbert F Smit[2]

Departments of Nuclear Medicine[1], Pulmonology[2] and Clinical Epidemiology and Biostatistics[4], University Hospital Vrije Universiteit, Amsterdam, The Netherlands.
Comprehensive Cancer Center Amsterdam[3], Amsterdam, The Netherlands

Thorax 2003;58:47-51

Abstract

Background:
A study was undertaken to investigate the effect of ^{18}F-fluorodeoxyglucose (^{18}FDG) positron emission tomography (PET) on diagnosis and management of clinically problematic patients with suspected non-small cell lung cancer (NSCLC).

Methods:
A prospective before-after study in a cohort of all 164 patients (university / community settings) referred for PET between August 1997 and July 1999. PET was restricted to cases where non-invasive tests failed to solve clinical problems. The impact on diagnostic understanding and management was assessed using questionnaires (intended treatment without PET, actual treatment choice after PET, post hoc clinical assessment).

Results:
Diagnostic problems especially pertained to unclear radiological findings (n = 112; 63%), mediastinal staging (n = 36; 20%) and distant staging issues (n = 16; 9%). PET findings were validated by reviewing medical records. PET had a positive influence on diagnostic understanding in 84%. Improved diagnostic understanding solely based on PET was reported in 26%, according to referring physicians, PET resulted in beneficial change of therapy in 50%. Cancelled surgery was the most frequent therapy change after PET (35%).

Conclusion:
^{18}FDG PET applied as "add-on" technology in patients with these clinical problems appears to be a clinically useful tool, directly improving therapy choice in 25% of patients. The value of increased confidence induced by PET scanning requires further evaluation.

Introduction

Medical imaging technology is rapidly expanding and the role of each modality is being redefined constantly. Positron emission tomography (PET) using ^{18}F-fluorodeoxyglucose (^{18}FDG) has emerged as an accurate imaging modality in patients with lung cancer.[1-3] Potential clinical indications include the differential diagnosis of benign versus malignant disease, initial (preoperative) staging, evaluation of suspected recurrences, and follow up after treatment. The use of PET in clinical practice is based predominantly on studies of technical performance and diagnostic accuracy.[4,5] To ensure an appropriate use of PET, such studies should be followed by an analysis of the impact of PET on management decisions, outcomes of care, and cost-effectiveness.

In the northwestern part of the Netherlands where this study was performed, a single PET scanner serves 2.7 million inhabitants, with 50% of its time slots available for clinical purposes. To restrict the use of PET to those patients that may benefit most, a program has been developed to evaluate the clinical usefulness of PET, investigating the cost-effectiveness of performing PET on a routine basis in the preoperative staging of non-small cell lung cancer (NSCLC)[6] and its impact as an "add on" technique in specific problem cases. To measure the clinical value of PET in the latter group, we performed a prospective before-after study in a cohort of clinically problematic cases, typically after an extensive conventional work-up. This study design was used during the early studies of computed tomographic (CT) scanning by Wittenberg et al[7] and allows a systematic assessment of the impact of a test on diagnostic understanding as well as on patient management within the clinical context.[8]

Methods

To be eligible for PET scanning, patients had to have suspected or proven NSCLC with a diagnostic problem which, according to the referring physician, could not be solved by conventional methods alone and in which the PET result might affect patient management. In an attempt to restrict PET scanning to such cases, referrals were only accepted after discussion of the case between this physician and the staff nuclear medicine physician in charge at the Clinical PET Centre of the VU University Medical Centre. PET scanning therefore typically followed an extensive conventional work-up. All patients routinely underwent laboratory tests, bronchoscopy, chest radiography and CT scanning extending from the neck to the upper abdomen (including liver and adrenal glands). Additional diagnostic tests were performed in cases with signs and symptoms suggestive of distant metastatic disease. Patients entered in randomized[9] or response monitoring trials[10] were not included in the present report.

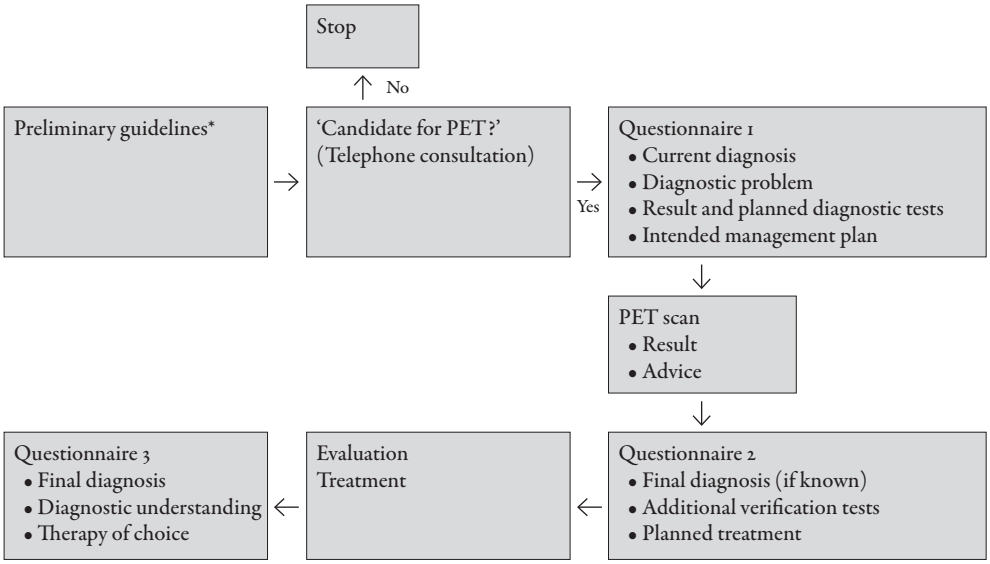


FIGURE 1. Study protocol.*
Suspected NSCLC, diagnostic problem insoluble by conventional imaging, potential impact on patient management.

Assessment of clinical value.

The impact of PET on diagnostic understanding, and therapy choice was investigated using three questionnaires (figure 1). These questionnaires were to be completed by the referring physician before PET scanning, shortly after PET scanning, and about 6 months after PET scanning, respectively. In the first questionnaire, information was requested regarding the histological diagnosis (if known), a definition of the current diagnostic problem, a differential diagnostic consideration, the results of diagnostic tests already performed and any planned diagnostic tests. In addition, the referring physician was requested to outline the intended patient management plan if PET scanning was not available. The second questionnaire requested information regarding the working diagnosis and planned treatment after PET scanning in addition to any diagnostic tests that had been ordered as a direct consequence of the PET scan result. In the final questionnaire, the referring physician was requested to convey the final diagnosis and to rate the overall usefulness of PET separately in terms of diagnostic understanding and therapy of choice according to the method of Wittenberg et al.[7] This method involves using a 5 point ordinal scale (box 1), with higher scores representing an increasing positive impact.

All questionnaires were checked for internal consistency between the pre-PET intentional management (questionnaire 1) and post-PET actual management (questionnaire 3). In the case of inconsistencies, the referring physicians were asked to review the cases in question and to revise the overall clinical value rating accordingly and these data were used in the analysis. In the case of PET negative – that is, suspected benign – coin lesions, follow-up was extended beyond 6 months by examining the medical records of these patients.

Diagnostic understanding (DU)

D = 1: PET confused my understanding of this patient's disease and led to investigations I would not otherwise have done

D = 2: PET confused my understanding of this patient's disease but did not lead to any additional investigations

D = 3: PET had little or no effect on my understanding of this patient's disease

D = 4: PET provided information which substantially improved my understanding of this patient's disease

D = 5: My understanding of this patient's disease depended upon diagnostic information provided only by PET (unavailable from any other non-surgical procedure)

Treatment choice (TC)

T = 1: PET led me to choose treatment which in retrospect was not in the best interests of the patient

T = 2: PET was of no influence in my choice of treatment

T = 3: PET did not alter my choice of treatment but did increase my confidence in the choice

T = 4: PET contribute to a change in my chosen treatment but other factors (other imaging tests, other diagnostic tests, changes in patient status) were equally or more important

T = 5: PET was very important compared with other factors in leading to a beneficial change in treatment

Box 1. Questionnaire on evaluation of PET impact

Management changes

Treatment (management) changes were considered "major" if treatment changed from one modality to another – for example, from medical to surgical/radiation/ no treatment or vice versa[11] – and "minor" if treatment changed within a modality – for example, altered medical, surgical or radiotherapy approach.

PET imaging

Whole body ^{18}FDG PET scans were performed with a dedicated PET scanner (ECAT EXACT HR+, CTI/Siemens). Emission scans, typically extending from mid-skull to mid-femur, were acquired in 2D mode, approximately 60 minutes after intravenous injection of 370 MBq (10 mCi) ^{18}FDG. Patients were asked to fast for at least 6 hours prior to the PET study. Oral intake of water was encouraged.

PET scans were corrected for decay, scatter and randoms. Scans were reconstructed as 128x128 matrices using filtered back projection with a Hanning filter (cut-off 0.5 cycles/pixel) resulting in a transaxial spatial resolution of 7 mm at full width half maximum. If possible, CT scan data were used for more precise anatomical localization of PET abnormalities suspected as being malignant. Referring physicians were informed by telephone of the result of the PET scan and an advice to the next step. Clinicians were urged to verify clinically decisive PET findings by conventional means (histology, imaging, follow-up) and to ignore unconfirmed hot spots. PET findings were retrospectively validated by examination of the medical records of the included patients. Histopathology and clinical follow up findings that showed a benign or malignant course were considered as a valid reference test.

Statistical analysis

Differences in diagnostic understanding or treatment choice between the three indications were tested by means of a two sided Kruskal-Wallis test. Wilcoxon-Mann-Whitney test was used to test differences between two samples. Changes in treatment plans before and after PET were tested by the marginal homogeneity test.[12]

Results

During a 23-month inclusion period, 179 patients with suspected NSCLC were referred for PET scanning. The referring physicians included pulmonologists (76%), oncologists (7%), internists (6%), radiotherapists (6%), neurologists (3%) and surgeons (1%) from 21 different university and community hospitals. Questionnaires were returned from 178 (99%) patients and a fully completed questionnaires (all questions answered) was obtained for 136 (76%) patients. Specifically, questionnaire 1 was fully completed for 83% of the patients, questionnaire 2 for 92%, and questionnaire 3 for 98%. Indications for PET could be subdivided in six groups: unclear radiological abnormality (including solitary pulmonary nodules and lung masses, n = 112; 63%), staging of the mediastinum (n = 36; 20%), distant staging issues (n = 16; 9%), response monitoring (n = 5; 2.8%), suspected recurrence (n = 5; 2.8%), and unknown primary (n = 5; 2.8%). The present report focuses on the first three clinical indications.

In these 164 patients, the clinical work-up before PET included laboratory tests, chest radiography, CT scan of the chest (including liver and adrenal glands) and bronchoscopy.

In patients with distant staging problems (n = 16) the work-up before PET consisted of bone scintigraphy and radiographic studies in the three patients with clinical concerns about skeletal metastases; CT evaluation of the abdomen typically preceded referrals with suspect adrenal enlargement or liver lesions in which biopsy was considered not feasible or had been inconclusive. In two patients in which chest CT scan had shown additional and indeterminate pulmonary lesions, bronchoscopic examination had been negative and it was not considered feasible to take biopsy specimens. In five patients with potentially solitary brain metastases, dissemination tests had included CT scanning (brain, chest, liver and adrenal glands) and bone scintigraphy. In general, the work-up of patients with unclear radiological findings before PET scanning conformed to national guidelines.[13]

The diagnostic problems concerning mediastinal staging leading to referral for PET (instead of invasive mediastinal staging) included former mediastinoscopy, thoracotomy or radiotherapy, indeterminate invasive staging results, medical inoperability, and "to determine the most appropriate surgical approach". After careful evaluation we were unable to identify a specific reason for choosing PET scanning as opposed to mediastinoscopy to determine mediastinal lymph node involvement in 10 patients.

In 29 out of the 179 patients the initially formulated management plans (to be carried out if PET had not been available) were not consistent with the final assessment of the impact of PET. For example,

the physician's written plan before PET was to perform a thoracotomy, and a thoracotomy was indeed performed but treatment choice was rated as 5 (PET was very important compared with other factors leading to a beneficial change in treatment). Such inconsistent assessments were revised by the referring physicians (specifically with respect to the questionnaire 3), and corrected in 28 cases.

	DU = 1	DU = 2	DU = 3	DU = 4	DU = 5	Missing	Total
Radiological abnormality	3	6	12	61	29	1	112
Mediastinal staging	1	1	1	21	11	1	36
Distant staging	0	0	2	10	2	2	16
Overall	4	7	15	92	42	4	164

Table 1. The impact of PET on diagnostic understanding (DU) ratings (defined in box 1)

Diagnostic understanding

The impact of PET on diagnostic understanding was analyzed for each clinical indication (table 1). Overall, PET was solely responsible for improved diagnostic understanding (DU = 5) in 26% (95% CI 19 to 33) of the patients and substantially contributed to diagnostic understanding (DU = 4) in 58% (95% CI 50 to 65). The effect of the PET result on diagnostic understanding was confusing and led to additional tests (DU = 1) in 3% (95% CI 1 to 6), and had no or little effect (DU = 3) in 9% (95% CI 5 to 15). The impact of PET on diagnostic understanding was not significantly different for the three clinical indications (p = 0.45). There was no significant difference (p = 0.85) in diagnostic understanding ratings between PET scans indicating malignancy where the tumor was finally proven to be malignant (true positives) and scans indicating benign disease where the lesion proved to be benign (true negatives). To evaluate the presence of a potential clinical learning curve of incorporating PET scanning results, we compared the diagnostic understanding ratings of "early" patients (the first five patients) referred by a particular physician to the ratings of later patients (the sixth and subsequent patients). The ratings in later patients tended to be significantly higher (p = 0.0192).

Diagnostic accuracy

Of the patients referred to resolve unclear radiological findings, 76 patients had a positive PET scan result which proved to be true positive in 68 patients (89%). Thirty six patients had a negative scan reading –that is, no focally enhanced FDG uptake suspicious for malignancy–which proved to be correct (true negative) in 34 patients (94%) either by "wait and see" policy (n = 32) or surgery (n = 2). The mean duration of follow up in these patients was 20 months (range 6-36). In two patients the PET scans proved to be false negative. These false negative cases included a patient with pulmonary fibrous tumor (the patient underwent a curative pneumonectomy) and a patient with mantle cell lymphoma (diagnosed 1 year after the PET scan). In one patient the indeterminate solitary pulmonary nodule proved to be true positive at surgery but PET was found to have missed micrometastatic involvement of mediastinal lymph nodes.

Of the patients referred for mediastinal staging, 24 patients had a positive PET scan result of which 22 were proven to be true positive as shown by pathology in 16 patients and by follow up in six patients; one was proven to be false positive (as shown by pathology) and one patient was lost to follow up. Eleven patients had negative scan results which were found to be true negative in 10 patients (as shown by pathology in six patients and by follow up in four: mean time from PET to last chest radiograph or CT scan was 15 months, range 13-17). In one patient the PET scan was found to be false negative (as shown by pathology). In one patient the scan trajectory did not include the mediastinum due to claustrophobia.

Of the patients referred because of distant staging issues, 10 were found to be true positive (as shown by pathology in six patients, follow up in two, and radiology in two). Six patients proved to have a true negative PET scan as shown by follow up in five patients (mean time of follow up 6 months, range 6-6). In one patient the PET result proved to be false negative (bone metastases).

Treatment change	No. of patients
Surgery to	
• Radiotherapy	6
• Chemotherapy	11
• Observation	18
Radiotherapy to	
• Surgery	1
• Chemotherapy	2
• Observation	3
Chemotherapy to	
• Surgery	2
• Radiotherapy	0
• Observation	2
Observation to	
• Surgery	3
• Radiotherapy	4
• Chemotherapy	1
Minor changes within	
• Surgery	14
• Radiotherapy	9
• Chemotherapy	2

Table 2. Treatment changes after PET (T = 4/5, n = 78)

Management changes

In 162 of the 164 cases studied explicit provisional therapeutic plans had been stated before PET. In 103 patients this involved surgery. After PET, surgery was the treatment most commonly abandoned (table 2). PET contributed to a decision to forego surgical treatment in 36 patients (35%; 95% CI 26 to 45) in whom it had been provisionally planned. Of the patients in whom surgery was not the proposed treatment before PET (n = 59), seven patients subsequently underwent surgery. In these patients the intended therapy had been observation in four patients, chemotherapy in two

patients, and radiotherapy in one patient. There was a significant change in terms of the "impact" of treatment for the patient, mainly toward a less aggressive approach (surgery→chemo-/radiotherapy→observation; p = 0.0001). The impact of PET on treatment was divided into major or minor changes as outlined previously. PET was responsible for changes of choice of treatment that were major in 55 patients (66%; 95% CI 55 to 76) and minor in 28 patients (34%; 95% CI 24 to 45).

	TC = 1	TC = 2	TC = 3	TC = 4	TC = 5	Missing	Total
Radiological abnormality	1	16	42	21	30	2	112
Mediastinal staging		3	11	10	10	2	36
Distant staging		3	4	3	5	1	16
Overall	1	22	57	34	45	5	164

Table 3. The impact of PET on patient management and its clinical assessment (treatment choice (TC) ratings as defined in box 1

Post hoc evaluation of treatment choice

The impact of PET on treatment choice was analyzed for each scan indication (table 3). According to the attending physician, PET was the most important factor leading to a beneficial change of treatment (TC = 5) in 45 of 159 patients (28%; 95% CI 21 to 35) patients and contributed to such change (TC = 4) in 34 (21%; 95% CI 15 to 28).

Of the 134 cases in which the physician reported increased diagnostic understanding, therapeutic plans remained unchanged in 59 cases (44%). No significant differences in changes of treatment choice for the three different indications were found (p = 0.65). Treatment choice ratings after PET scanning indicating malignancy when the suspected lesion was indeed found to be malignant were not different from scans indicating a benign lesion found to be benign (p = 0.27). Like diagnostic understanding, the treatment choice ratings were significantly higher for later patients than for early patients (p = 0.037).

Discussion

A new test that appears to be more accurate than the standard ones will generate a clinical demand, even if it's effect on clinical outcome measures is still unclear. With scarce technology like PET overconsumption may result precluding general accessibility. Evidence-based guidelines for routine use are therefore needed, so that the available scanning capacity can be adjusted to the expected demand. However, guidelines aim at the average patient and may not be applicable in specific situations. In this prospective, multicenter before-after study the reported clinical impact of ^{18}FDG PET as an "add-on" technology to solve diagnostic problems in patients with suspected NSCLC was considerable. Clinical compliance with the PET results was high, and PET was reported to have led to beneficial management changes (TC≥4) in 50% of the patients in the three clinical situations

investigated. In addition, a positive influence on diagnostic understanding (DU≥4) by PET was observed in 84% of the patients. Put in a more conservative way, PET proved to be the key diagnostic tool in one of every four patients referred for PET (DU/TC = 5).

Interestingly, we observed an increasing appreciation of PET over time. Even though other explanations may also be valid, individual consultation and feedback as done in our setting, is known to improve patient referral patterns.[14]

Interpretation of the classification of "important contribution" to treatment choice by PET (TC = 4) is not straightforward. It is recognized that, in most clinical situations, decisions are made on the basis of a number of factors. Patient management depends on the preoperative assessment of the probability of disease, which is a joint function of multiple diagnostic indicators such as signs, symptoms and test results together with the effectiveness of the invasive procedures that follow them. This complicates the assessment of the contribution of a single test to a change in patient management. Even though the phrasing of the "contributive" ratings (DU/TC = 4) may benefit from accentuation, such positive perceptions may always contain a spectrum of clinical relevance which is difficult to translate into outcome measures. The assessment of the true value of "contributive" rather than directly decisive PET findings (TC = 4 v TC = 5) is therefore best done in a randomized study design.

Some studies have recently addressed the clinical impact of PET. The methodologies and patient spectra were variable, but the reported management changes (65-70%)[15-17] are uniformly higher than those observed as a by-product in accuracy studies (10-59%).[18-19] This underlines the fact that management change is multifactorial and does not merely depend on a single test (such as PET).

Alternatively, "clinical value" studies may have overestimated the true clinical contribution of PET. Firstly, the clinical impact of a new technology depends on the quality of the previous clinical work-up; poorly performed conventional staging before PET scanning would overestimate its actual value. We therefore made an effort to restrict PET referrals to cases in which conventional investigations had indeed been performed and had failed. As we have shown, this was the case in the majority patients. Further, a retrospective analysis of the pre-PET work-up showed adherence to internationally accepted guidelines in the majority of patients. Secondly, whether a specific test contributed significantly is a matter of judgment, and thus subject to disagreement, error and imprecise measurement.[8] This was, indeed, the case in our study; inconsistencies were identified in 18% of the questionnaire responses. To strengthen the evidence of before-after studies, independent reviewing of the data by experts has been suggested. This has been shown to reduce the presumed benefit of a new technology as assessed with this type of study design.[20] However, such findings may also reflect the heterogeneity of daily clinical practice in which patients are actually diagnosed and treated. Thirdly, unconscious bias of the referring clinicians in favor of the new technology may have affected the results. We cannot rule out that this has occurred but the opposite may also be true. Even though the sample was not randomly chosen, we found no such effect in the medical records of the cases in which a prolonged follow up was needed and the data were derived from a broad spectrum of hospitals.

The questionnaires used do confirm a distinction between the clinical impact of a test on diagnostic understanding, patient management, and (retrospective) clinical assessment of the appropriateness

of these changes. The data clearly show that the perceived benefit of PET scanning consists of altered patient management but, to an even greater extent, of increased diagnostic understanding or confidence in cases where patient management was not altered. In their present form, the questionnaires do not allow estimation of the amount of clinical uncertainty. In our opinion, studies such as this may serve to estimate the relative merits of PET for different indications within a specific clinical context. If PET fails to show clinical impact, the presumed indication for PET may be removed from the list, whereas promising results warrant further investigation. Our data do not represent consecutive patients presenting with a similar clinical problem, and as such, our results cannot be extrapolated to imply the routine use of PET in all patients with suspected NSCLC. Estimation of the cost-benefit of such an application requires a direct comparison between patients subjected to PET and conventional work-up. Such a study is currently ongoing in the Netherlands.

In summary, controlled implementation of PET, as a 'last resort' diagnostic modality, improved patient management in at least 25% of clinically problematic cases with suspected NSCLC. The combination of preliminary guidelines, intensive feedback, and prospective monitoring may promote the effective use of scarce technology.

Acknowledgement

The authors thank A. Kalwij and C. Karga (secretaries, Clinical PET Centre) for collecting all the questionnaires.

References

1. Dwamena BA, Sonnad SS, Angobaldo JO et al. Metastases from non-small cell lung cancer: mediastinal staging in the 1990s – meta-analytic comparison of PET and CT. *Radiology* 1999; 213:530-536.
2. Gould MK, Maclean CC, Kuschner WG et al. Accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions: a meta-analysis. *JAMA* 2001; 285:914-924.
3. Pieterman RM, van Putten JW, Meuzelaar JJ et al. Preoperative staging of non-small-cell lung cancer with positron- emission tomography [comment]. *N Engl J Med* 2000; 343:254-261.
4. Adams A, Flynn K. Positron Emission Tomography – descriptive analysis of experience with PET in VA. *Technology Assessment Program* 10, i-A5-4. 1998. Boston, USA.
5. Commonwealth department of health and aged care. Report of the commonwealth review of positron emission tomography. Health access and financing division, Australia, 2000. http://www.health.gov.au/haf/msac
6. Van Tinteren H, Hoekstra OS, Smit EF et al.. Effectiveness of positron emission tomography in the preoperative assessment of patients with suspected non-small-cell lung cancer: the PLUS multicentre randomised trial. *Lancet* 2002;359:1388-93.
7. Wittenberg J, Fineberg HV, Black EB et al. Clinical efficacy of computed body tomography. *AJR Am J Roentgenol* 1978;13:5-14.
8. Guyatt GH, Tugwell PX, Feeny DH et al. The role of before-after studies of therapeutic impact in the evaluation of diagnostic technologies. *J Chronic Dis* 1986;39:295-304.
9. Van Tinteren H, Hoekstra O, Smit E et al. on behalf of the IKA-PLUS Study Group. Towards less Futile Surgery in Non Small Cell Lung Cancer? A Randomized Clinical Trial to Evaluate the Cost-effectivensess of Positron Emission Tomography. *Controlled Clinical Trials* 2001; 22:89-98.
10. Vansteenkiste JF, Stroobants SG, Hoekstra C et al. ^{18}Fluorodeoxyglucose -2-Deoxyglucose Positron Emission Tomography (PET) in the Assessment of Induction chemotherapy (IC) in Stage IIIa0N2 NSCLC: a Multi-Center Prospective Study [abstract]. *J Clin Oncology* 2001;20,313a.
11. Seltzer MA, Valk PE, Wong CS et al. Prospective survey of referring physicians to determine the impact of whole body FDG-PET on management of cancer patients [abstract]. *J Nucl Med* 2000;428.
12. Agresti A. Categorial Data Analysis. Ed. John Wiley & Sons, New York, 1990.
13. van Zandwijk N. [Consensus conference on the diagnosis of lung carcinoma (see comments)]. *Ned Tijdschr Geneeskd* 1991;135:1915-1919.
14. Eccles M, Steen N, Grimshaw J et al. Effect of audit and feedback, and reminder messages on primary-care radiology referrals: a randomised trial. *Lancet* 2001;357:1406-1409.
15. Kalff V, Hicks RJ, MacManus P et al. Clinical Impact of (18)F Fluorodeoxyglucose Positron Emission Tomography in Patients With Non-Small-Cell Lung Cancer: A Prospective Study. *J Clin Oncol* 2001;19:111-118.
16. McCain TW, Dunagan DP, Chin R et al. The usefulness of positron emission tomography in evaluating patients for pulmonary malignancies. *Chest* 2000;118:1610-1615.
17. Tucker R, Coel M, Ko J et al. Impact of fluorine-18 fluorodeoxyglucose positron emission tomography on patient management: first year's experience in a clinical center. *J Clin Oncol* 2001;19:2504-2508.
18. Gambhir SS, Czernin J, Schwimmer J et al. A tabulated summary of the FDG PET literature. *J Nucl Med* 2001;42:4S-8S.
19. Saunders CAB, Dussek JE, O'Doherty MJ et al. Evaluation of fluorine-18-Fluorodeoxyglucose Whole Body Positron Emission Tomography Imaging in the Staging of Lung Cancer. *Ann Thorac Surg* 1999;67:790-7.
20. Goldman L, Feinstein AR, Batsford WP et al. Ordering patterns and clinical impact of cardiovascular nuclear medicine procedures. *Circulation* 1980;62:680-687.

CHAPTER 3

Diagnostic imaging randomized controlled trials: a review

Harm van Tinteren[1], Otto S Hoekstra[2,3], Maarten Boers[3]

Comprehensive Cancer Center Amsterdam[1], Amsterdam, The Netherlands.
Departments of Nuclear Medicine & PET Research[2] and Clinical Epidemiology & Biostatistics[3],
VU University Medical Center, Amsterdam, The Netherlands.

Submitted to *Journal of Clinical Epidemiology*

Abstract

Objective

To evaluate the impact of a new diagnostic imaging device on patient outcome, a randomized controlled trial (D-RCT) is considered to provide the best evidence. We reviewed the literature for D-RCTs, for time trends and study characteristics, applying a taxonomy accounting for main contrasts and outcome measures.

Study design and Methods

Using Pubmed, a search on general D-RCTs included three distinctive years of publication, between 1990 and 2002. Another search focused on D-RCTs between 1990 and 2005 in which Magnetic Resonance Imaging (MRI) played a major role in the comparison.

Results

The majority of studies pertained to testing different tracers or acquisition parameters with one particular device. An increase over time was observed in studies comparing devices over studies testing tracers. Only 15 D-RCTs were identified in which patient outcomes were studied.
The search on MRI revealed 13 D-RCTs of which 8 were published after 2000. In three studies, two devices were tested in the same patient, while in the others patients were randomly allocated to one of two strategies.

Conclusion

Although generally advocated, randomized controlled trials in diagnostic imaging are not frequently performed. As a consequence, evidence on (cost)-effective use of imaging tests is lacking in general. A positive trend in D-RCTs is observed.

Introduction

In general, establishing effectiveness of diagnostic imaging devices has two components: [1] establishing the accuracy of the test and [2] establishing its clinical value. In the ideal world, an image is pathognomonic for a specific diagnosis, and patient management follows from this diagnosis. In the reality of imaging, there are "gray areas", doubts, conjectures.[3] This is especially pressing in settings where tests are used sequentially (e.g. to stage a cancer patient) followed by invasive procedures with variable yield. The use of diagnostic devices without defined benefits may cause harm to patients, including [1] further unnecessary diagnostic testing, including invasive processes [2] directing patients to inappropriate therapy, and [3] creating unwarranted patient anxiety from abnormal test results. An application for marketing authorization for a diagnostic agent/test[4] should therefore address both accuracy and clinical value as proposed by Fryback and others in a hierarchical framework.[5] In this often cited hierarchical framework, the three last steps support the clinical efficacy: patient management efficacy, patient outcome efficacy and societal efficacy. Especially for the impact on patient outcome, a randomized controlled trial (RCT) is usually considered to provide the best evidence, just like in medical intervention research.[5-8]

In 1996 we were asked to plan a study assessing the cost-effectiveness of PET in NSCLC. Because we considered a RCTs to be the optimal in this situation, we performed a literature search for randomized diagnostic imaging studies, to help and design the study. We identified many papers applying randomization, but very few appeared to address patient outcomes as a function of the applied diagnostic test(s) which was in fact our main research question. In 2005 we updated the literature search with another 10 years.

The aim of the present work was first to design and apply a taxonomy of diagnostic RCT's (D-RCT's) accounting for main study contrasts and outcome measures, and second to investigate trends in time with respect to D-RCT's addressing patient outcomes as a function of the imaging test.

Methods

Literature search

Two complementary searches in MEDLINE® were performed. Both search strategies included terms suggestive of randomized controlled trials as proposed by Dickersin.[9] The set of keywords were optimized interactively with experience from known diagnostic studies and by investigating the results. We included studies on patients referred for diagnostic work-up for some suspected state of disease or for a condition that required medical intervention. Screening trials, where the issue was to screen or not to screen in 'non-diseased' populations were therefore excluded.

The abstracts of all identified articles were reviewed initially by one author (HvT), and classified according to the research topic and the design. When abstracts were not sufficiently informative, full papers were collected. Of all relevant abstracts, full articles were gathered for further review. No formal systematic complementary hand-search procedures were applied.

Our first search included three distinctive years of publication at intervals of 6 years to be able to capture possible developments in time: 1990, 1996, which was the year just before we actually designed our first D-RCT to study the role of FDG-PET in NSCLC[10] and 2002, the year of its publication.[11]

In a second approach we focused on D-RCTs in which Magnetic Resonance Imaging (MRI) played a major role in the comparison. MRI is the most recent tomographic imaging device before PET and may bear many similarities in terms of introduction into clinical practice. Moreover, by limiting the search in terms of a particular device it was feasible to include a longer time range as to study possible time trends. Randomized Controlled Trials and Magnetic Resonance Imaging were included as MESH terms together with a time range from January 1990 until October 2005 (Table 1b). Again, mass screening was excluded.

Taxonomy

We designed a taxonomy based on a classification of the main study contrast as well as the outcome measures (Table 2). The classification of endpoints was derived from the hierarchical model of Fryback.[5] In this model, the first level is concerned with technical efficacy, including technical aspects of the image and safety aspects of the imaging contrast agent or radiopharmaceutical. The second level involves the diagnostic accuracy, often expressed in measures of sensitivity and specificity. In this study, the third and fourth level, associated respectively with diagnostic thinking (impact) and therapeutic impact were taken together because they are usually studied together in each other's extension in a particular study design. The fifth level concerns the patient outcome efficacy and a sixth level is addressed if costs and benefits from a societal viewpoint are included. For each study, the highest level of endpoint reported was assigned although parameters of previous levels were often included in the results.

We extracted a set of predefined parameters of individual studies: the contrast, defined as the actual comparison studied, endpoints, disease type, patient spectrum and the timing of randomization, e.g. before or after the use of the device or tracer. This accounts for the issue whether patients were randomized to a particular device or whether scans were randomized among observers. Finally, we extracted population characteristics and logistical aspects to assess the external validity, such as the number of patients and whether the study was mono- or multicentric.

```
((("Diagnostic Imaging"[MAJR] OR "Signs and Symptoms/radiography"[MAJR]) OR "Signs and
Symptoms/radionuclide imaging"[MAJR]) OR "Signs and Symptoms/ultrasonography"[MAJR]) AND
((((("clinical trials"[Mesh:noexp] AND (((randomized[tiab] OR randomised[tiab]) OR control[ti]) OR
controlled[tiab])) OR ("controlled clinical trials"[mesh] AND (randomized[tiab] OR randomised[tiab])))
OR "Randomized Controlled Trial"[Publication Type]) OR ("Controlled Clinical Trial"[Publication Type]
AND (randomised[tiab] OR randomized[tiab])))
```

Table 1a. Medline search 1 (see text)

```
"Magnetic Resonance Imaging"[MAJR]AND ("randomized controlled trial"[Publication Type] OR
"randomized controlled trials"[MESH Terms]) AND ("1990/01/01"[PDAT] : "2005/10/31"[PDAT])
```

Table 1b. Medline search 2 (see text)

Results

The initial search (Table 1a) revealed 66, 194 and 243 potential titles in 1990, 1996 and 2002. Of these, respectively 30 (45%), 70 (36%) and 50 studies (21%) were eligible because they used a design that involved some type of randomization of the diagnostic technique or an aspect related to diagnosis. In 27 studies diagnostic devices or strategies were compared (table 2, categories I, II), while the majority of studies (82%) pertained to testing different tracers or acquisition parameters with one particular device (table 2, III-VII). A substantial increase over time was observed in the number of studies comparing devices in contrast to studies testing tracers (table 2).

		Year of publication						All	
		1990		1996		2002			
		N	(%)	N	(%)	N	(%)	N	(%)
	Contrast								
I	Imaging Device A versus B	.	.	6	(8.6)	8	(16.0)	14	(9.3)
II	Imaging Device A versus no A	3	(10.0)	1	(1.4)	9	(18.0)	13	(8.7)
III	Different acquisition parameters, single device	2	(6.7)	13	(18.6)	12	(24.0)	27	(18.0)
IV	Tracer A versus B	16	(53.3)	25	(35.7)	4	(8.0)	45	(30.0)
V	Tracer A: different concentrations	2	(6.7)	5	(7.1)	7	(14.0)	14	(9.3)
VI	Different disease status, single device	.	.	1	(1.4)	.	.	1	(0.7)
VII	Devices or interventions supporting the imaging	7	(23.3)	19	(27.1)	10	(20.0)	36	(24.0)
	All	30	(100.0)	70	(100.0)	50	(100.0)	150	(100.0)

Table 2. D-RCT's categorized by type of contrast by year of publication

Fourteen studies contrasted one device directly against another (table 2, I). We were able to distinguish different types of designs. For example, in some studies two devices were applied and compared within the same patient. Randomization either served to account for disease progression between tests (randomization of the order in which the tests were performed) or to reduce case and observer bias if added value of tests was at stake (randomization of the order in which test results were offered to observers). An example of the first situation is the comparison of diffusion-weighted MRI with CT in hyperacute stroke patients.[12] The second situation was found in a study on the follow-up of patients surviving aortic dissection, where transesophageal echocardiography was contrasted to X-ray computed tomography.[13] There, the additive value of one test over the other was studied in the context of diagnostic impact. In some studies with two devices applied in the same patient, the or-

der was not important, and only the scans were randomized and interpreted retrospectively by two observers. This situation was found in a study in patients with anterior mediastinal masses where the objective was to compare chest radiography with computed tomography in the prediction of a specific diagnosis.[14] Overall, such type of design, where randomization referred to the 'shuffling' of images after scanning and before interpretation by (blinded) reviewers, was applied in eleven out of 150 identified randomized studies. In only two of the 14 studies within the category of device comparisons (Table 2, I), patient outcomes were part of the endpoint of investigation. One looked at the value of computer-aided diagnosis versus contrast radiography on morbidity and mortality of patients with acute small bowel obstruction[15] and another at the impact of MRI versus plain radiographs for low back pain on patient outcomes.[16]

In 13 studies patients were randomized to a (conventional) diagnostic strategy or to the same strategy with a new test on top of it (table 2, II). All focused on an endpoint beyond accuracy. The spectrum of diseases studied included cancer, cardiovascular- and musculoskeletal problems (e.g. low back pain). Sample sizes ranged from 15 patients in a study to validate MRI versus radio-opaque markers for diagnosis of delayed gastric emptying in diabetic patients, to 2475 in a trial comparing a conventional strategy in emergency department patients versus the usual strategy supplemented with results from acute resting myocardial perfusion imaging using single-photon emission computed tomography.

The largest group of studies (table 2, IV,V) used randomization for comparison of different imaging agents (radiological contrast agents or radiopharmaceuticals) using the same scanner (49%). The number of such studies decreased over time representing 60% of all eligible studies in 1990, 43% in 1996 and 22% in 2002 (table 2, IV). Studies investigating devices and interventions supporting the imaging process, (eg. standard versus pediatric probes for transesophageal echocardiography, or different sedatives) were also frequently reported and their number was stable over time (table 2, VII). For the category of testing different agents or acquisition parameters with one particular test device, typically the objectives or endpoints are related to diagnostic accuracy and/or technical efficacy including safety parameters and sometimes to associated costs (table 3). In the case of supportive interventions, comforting the situation of the patient and improving patient recovery (patient outcome or costs) were considered relevant endpoints, but with little relevance to the potential of the diagnostic device itself. Therefore within this taxonomy, the endpoints were graded as technical or improving diagnostic efficacy at most.

	Endpoints					
	Technical performance	Diagnostic efficacy	Diagnostic/ therapeutic impact	Patient outcome	Cost-effectiveness	All
	N	N	N	N	N	N
Contrast						
Imaging Device A versus B	3	5	4	2	.	14
Imaging Device A versus no A	.	1	1	7	4	13
Different acquisition parameters, single device	7	19	1		.	27
Tracer A versus B	22	21	2	.	.	45
Tracer A: diff concentrations	8	6	.	.	.	14
Different disease status, single device		1				1
Devices or interventions supporting the imaging	32	4				36
All	75	60	6	14	4	150

Table 3. D-RCTs categorized by type of contrast and endpoint of the study

MRI-studies

Of the initially identified 503 MRI papers, about one-third (156) reported the results of an actual RCT, focusing on an aspect of diagnosis. Of these 156 studies, the majority (76%) pertained to the analysis of different acquisition parameters (fast-spin echo versus conventional spin echo sequences, or imaging at 0.5T versus 1.5T) or contrast agents in a particular setting. Contrast agents or more frequently, their optimal concentration were investigated in 77 studies (49%). In 9 studies, MRI's were made both in patients known to have a specific disease and controls known to be without that disease. An example of such a study was a trial where MR studies were performed in 100 subjects with clinical histories of stroke and 203 subjects without reported histories of stroke. MR scans were independently evaluated by two trained neuroradiologists for the presence of small (≤ 3 mm) and large (> 3 mm) 'infarctlike' lesions.[17] Here, randomization pertained to the random distribution of images among multiple observers (see before (14)). In fourteen studies (9%), tools or agents (sedatives, enemas) to facilitate MRI were tested by means of an RCT.

Contrast	Outcome					Total
Frequency	Technical performance	Diagnostic efficacy	Diagnostic/ therapeutic impact	Patient outcome	Cost-effectiveness	
device A vs B	0	16	3	2	3	24
device A vs no A	1	2	2	4	4	13
Different acquisition parameters, single device	8	11	.	0	0	19
Tracer A vs B	10	18	.	0	0	28
Tracer A: different concentrations	21	17	.	0	0	38
Tracer A versus no A	5	6	.	0	0	11
Different disease status, single device	4	5	.	0	0	9
Devices or interventions supporting the imaging	13	1	.	0	0	14
Total	62	76	5	6	7	156

Table 4.

Eighteen publications were found that involved a comparison of MRI to alternative test(s) using endpoints beyond diagnostic accuracy. However, a more detailed analysis revealed that only 13 were unique D-RCT's and 5 other manuscripts referred to interim analyses or derived subgroups (Table 5a and b). In 10 of 13 D-RCT's, patients were randomly allocated either to a diagnostic process including MRI or to a device or diagnostic process without MRI. In the three remaining studies, patients underwent MRI as well as the competitive procedure, and the results were then randomized for use in diagnostic and therapeutic decisions. Musculoskeletal diseases were studied most frequently. Other individual studies involved recurrent breast cancer, gallstone pancreatitis, peripheral arteries, guidance of stereotactic procedures in Parkinson's disease and the decision to perform caesarean delivery with breech presentation at term. Usually, MRI was contrasted to a procedure without MRI on top of a conventional policy. One study investigated early imaging versus a policy without immediate imaging, where the choice of the imaging device, either CT or MRI, was left to the investigator.[18]

Major endpoints of the 13 studies are also reported in table 5a and b. Diagnostic and therapeutic impact was usually measured in terms of the need for and number of additional tests. Quality of life, QALY's and number of days immobilized were parameters of patient outcome. No study considered survival as major endpoint. Six of these 13 studies accrued less than 100 patients, five between 100 and 500 patients and two accrued 500 and 782 patients, respectively. No exceptional difficulties in patient accrual were reported. Of the 13 studies, 4 were published in 2005, 4 between 2000 and 2005 and 5 between 1990 and 2000.

In bovenstaande tekst twee maal 'Table 4a and b' gewijzigd in 'Table 5a and b' ???

Table 5a. Randomising patients to MRI or another device/process

Authors	Year of publication	Accrual	Devices	Disease	Main endpoint	Management protocolized	Sample size
Dixon AK, et al.	1993	?	MRI versus CT	Breast cancer	– Diagnostic efficacy – QoL (6 months)	Not specified	57
Blanchard TK, et al.	1999	12 months accrual	MRI versus arthrography	Shoulder problems	– Diagnostic/therapeutic impact – Additional imaging/surgical procedures	Not protocolized but measured before/after	53
Jarvik JG, et al.	1997	?	MRI versus plain radiographs	Low back pain	– Back-related disability – QoL – Costs	Not specified	62
Bryan S, et al.	2001	?	MRI versus no MRI	Knee joint	– avoidance of surgery (diagnostic/therapeutic) – QoL – Costs	Not specified	118
Jarvik JG, et al.	2003	Nov '98-June '00	MRI versus radiographs	Low back pain	– Back-related disability – QoL – Costs	Not specified	380
Gilbert FJ, et al.	2004	Nov '96-June '99	Early versus late selective imaging by MRI	Low back pain	– Back pain – QoL, QALY – Costs	Not specified	782
Brooks S, et al.	2005	?	MRI versus no MRI	Scaphoid fractures	– No of days immobilized – No of used health care units – Costs	Not specified	28
Hallal AH, et al.	2005	Feb '01-May '03	MRI versus no MRI	Gallstone pancreatitis	– Detecting choledocholithiasis	Protocolized	63
Ouwendijk R, et al.	2005	Dec '01-Sep '03	MRI versus multi-detector row CT angiography	Peripheral arterial	– Therapeutic confidence – Additional imaging – Ankle-brachial index – QoL	Not specified	157
Nikken JJ, et al.	2005	Aug '99-May '01	MRI versus no MRI	Acute peripheral joint injury	– N° additional tests – Duration of D-process – Days absent from work, QoL	Additional treatment not strictly specified	500

Table 5b. Randomised for further decision making after applying different strategies in the same patients

Authors		Reference	Devices	Disease	Main Endpoint	Management protocolized	Sample size
Blanchard TK, et al.	1999	—	MRI versus conventional arthrography	Shoulder problems	– Diagnostic impact – Therapeutic impact – Diagnostic performance	Not protocolized but measured before/after	104
Honey CR; Nugent RA;	2000	—	MRI vs CT	Parkinson's Disease	– Surgical outcome (stereotactic guidance)	Protocolized	24
van Loon AJ, et al.	1997	Jan '93-April '96	MRI vs no MRI	Breech presentation	– Elective and emergency caesarian-section rates – Early condition neonate	Procotolized	235

Discussion

Imaging tests are not introduced through carefully planned prospective evaluation of diagnostic and therapeutic impact and (cost)-effectiveness. In other words, the generally appreciated and advocated hierarchical framework of test evaluation is rarely fully applied to diagnostic imaging tests. This finding is in agreement with an analysis of diagnostic and screening radiology outcomes literature from 1990. Out of 4,205 articles potentially describing radiology outcomes investigations only 40 randomized controlled trials were found.[19] Also, a recent search in the Cochrane Central Register of Controlled Trials (issue 1, 2005) showed that only 4.2% of the records dealt with diagnostic tests or screening.[20] Our primary aim was to focus on studies where patients would be prospectively randomized to undergo some diagnostic test. Randomization is primarily used to balance known (and unknown) prognostic factors to prevent confounding the relation between intervention and outcome. If that is accomplished, the result of the imaging procedure is the only variable between the study groups.[21] In our search the randomization-label was often 'misused' to indicate a procedure of shuffling of images retrospectively obtained in order to distribute them 'randomly' to observers. Here, randomization is merely used to blind observers for any systematic order in the images. Obviously, such a design is unsuitable for studying patient outcomes.

Our general search for D-RCT's revealed 19 trials studying patient management, patient outcome or societal efficacies of imaging tests in 3 different years. No particular device or disease dominated. The search for studies randomizing patients to MRI versus another device or process revealed 13 primary studies in the past 15 years.

Remarkably, randomization seems to be totally accepted in the studies of different contrast agents or different concentrations of the same agent. In 1996 even a meta-analysis of 57 randomized double-blind clinical trials of one of the agents was published.[22] Paradoxically, randomization also appears to be an accepted and commonly used tool in screening trials (which often include the use of imaging test). Screening trials deal with the highest level of patient outcome (cost-effectiveness of screening versus no screening). In the earlier cited synopsis on radiology outcomes literature, the 40 randomized trials included 8 breast screening trials and 4 in lung cancer.[19] A possible explanation could be the fact that screening trials are no more the domain of a single center, but can only be successful in large areas and require the involvement of many more disciplines including methodologists, economists and health technology assessors.

In accordance to our first conclusion, we noted a lack of literature search strategies to identify RCT's on diagnostic tests as opposed to the situation with accuracy studies. In recent years several groups spent a credible effort in developing search strategies for diagnostic research and articles related to clinical guidelines and the appropriateness, process, outcome, cost and economics of health services.[23-27] Most of these search strategies aim to find diagnostic accuracy studies and include search terms like sensitivity and specificity. Nowadays, dedicated algorithms are available in Pubmed to facilitate the literature search.[23] This algorithm however, was not useful to identify diagnostic studies with objectives beyond the assessment of accuracy. More initiatives in this area are warranted.

The number of relevant studies detected by our first search is so low that it does not allow finding patterns or methods related to a type of diagnostic device or to a particular disease indication. However, over the three years, an increase in the number of relevant studies was observed.

Because we were primarily interested in outcome studies (both patient outcome and costs) related to a new diagnostic imaging device, bearing similarities to PET, we focused on studies investigating the role of MRI in the last decades. The MRI experience suggests that, more recently, individual centers set out to perform such RCT's but it is certainly not a widely accepted approach.[28] The majority of studies dealt with musculoskeletal problems or arterial disease and were mainly coming from specific groups from the UK[18,29-32] or the Netherlands.[33,34] Endpoints included the need for additional imaging or surgical procedures to quality of life and ultimately cost. All studies focused on the benefit of entire diagnostic strategies, one of them including MRI, rather than the assessment of MRI as a single test. In three studies, the conventional and the experimental devices were applied both in all patients. When the availability is not a concern and no particular risk is involved with the new device, this design may have the advantage of obtaining information with the technique in all patients. Often however, applying the new device is costly and the technique may not be sufficiently available for a new indication. In our experience, medical ethics committees dislike designs where all patients are exposed to an experimental device without using the information for clinical decision making.

Our search did not reveal any study in which patients were randomized to further testing or therapy on the basis of a test result. Such a design has been suggested to gain efficiency when the efficacy of the therapy or management following the test is uncertain.[8] Moreover, further management was only specified in 3 out of the 13 randomized MRI studies. Running trials without a protocol translating the test results to clinical management decisions has been criticized because of the fact that too many variables are at stake. Differences might be due to the test, management decisions guided by those tests or low effectiveness of subsequent treatment. The choice of the endpoint of diagnostic trials in this respect is very important and should be as close and relevantly linked to what may be expected as a result of the test.

None of the studies mentioned particular difficulties in ethical issues or accrual of patients. Most of the studies had an enrollment period of less than 3 years, which is totally acceptable and very similar to the situation in medicine trials. However, most trials were rather small, certainly compared to what is usually required to reliably assess the efficacy of new drugs. In general the papers lacked a proper justification of the sample size.

In conclusion, there seems to be a serious discrepancy between the often advocated principles for evaluation of new diagnostic technologies and devices and the methods that are actually applied in studies. As a consequence, evidence on (cost)-effective use of imaging tests is lacking in general. The experience with MR, corroborated by our own experience,[11,35] although limited to a few situations and centers, shows that they are feasible and recently more often performed.

Reference List

1. Flynn K, Adams E. Technology Assessment: Positron Emission Tomography. *HTA record 988699*, 1996.
2. Herder GJ, Verboom P, Smit EF et al. Practice, efficacy and cost of staging suspected non-small cell lung cancer: a retrospective study in two Dutch hospitals. *Thorax* 2002; 57(1):11-14.
3. Bruzzi JF. The words count – radiology and medical linguistics. *N Engl J Med* 2006; 354(7):665-667.
4. Committee for proprietary medicinal products (CPMP). Points to consider on the evaluation of diagnostic agents. 15-11-2001.
5. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991; 11(2):88-94.
6. Freedman LS. Evaluating and comparing imaging techniques: a review and classification of study designs. *Br J Radiol* 1987; 60(719):1071-1081.
7. Hunink MG, Krestin GP. Study design for concurrent development, assessment, and implementation of new diagnostic imaging technology. *Radiology* 2002; 222(3):604-614.
8. Bossuyt PM, Lijmer JG, Mol BW. Randomised comparisons of medical tests: sometimes invalid, not always efficient. *Lancet* 2000; 356(9244):1844-1847.
9. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994; 309(6964):1286-1291.
10. van Tinteren H, Hoekstra OS, Smit EF, Verboom P, Boers M. Toward less futile surgery in non-small cell lung cancer? A randomized clinical trial to evaluate the cost-effectiveness of positron emission tomography. *Control Clin Trials* 2001; 22(1):89-98.
11. van Tinteren H, Hoekstra OS, Smit EF et al. Effectiveness of positron emission tomography in the preoperative assessment of patients with suspected non-small-cell lung cancer: the PLUS multicentre randomised trial. *Lancet* 2002; 359(9315):1388-1393.
12. Fiebach JB, Schellinger PD, Jansen O et al. CT and diffusion-weighted MR imaging in randomized order: diffusion-weighted imaging results in higher accuracy and lower interrater variability in the diagnosis of hyperacute ischemic stroke. *Stroke* 2002; 33(9):2206-2210.
13. Maffei S, Baroni M, Terrazzi M et al. Ambulatory follow-up of aortic dissection: comparison between computed tomography and biplane transesophageal echocardiography. *Int J Card Imaging* 1996; 12(2):105-111.
14. Ahn JM, Lee KS, Goo JM, Song KS, Kim SJ, Im JG. Predicting the histology of anterior mediastinal masses: comparison of chest radiography and CT. *J Thorac Imaging* 1996; 11(4):265-271.
15. Bogusevicius A, Maleckas A, Pundzius J, Skaudickas D. Prospective randomised trial of computer-aided diagnosis and contrast radiography in acute small bowel obstruction. *Eur J Surg* 2002; 168(2):78-83.
16. Jarvik JG, Deyo RA, Koepsell TD. Screening magnetic resonance images versus plain films for low back pain: a randomized trial of effects on patient outcomes. *Acad Radiol* 1996; 3 Suppl 1:S28-31.:S28-S31.
17. Bryan RN, Manolio TA, Schertz LD et al. A method for using MR to evaluate the effects of cardiovascular disease on the brain: the cardiovascular health study. AJNR *Am J Neuroradiol* 1994; 15(9):1625-1633.
18. Gilbert FJ, Grant AM, Gillan MG et al. Low back pain: influence of early MR imaging or CT on treatment and outcome – multicenter randomized trial. *Radiology* 2004; 231(2):343-351.
19. Blackmore CC, Black WC, Jarvik JG, Langlotz CP. A critical synopsis of the diagnostic and screening radiology outcomes literature. *Acad Radiol* 1999; 6 Suppl 1:S8-18.:S8-18.
20. Gluud C, Gluud LL. Evidence based diagnostics. *BMJ* 2005; 330(7493):724-726.
21. van Tinteren H, Hoekstra OS, Boers M. Do we need randomised trials to evaluate diagnostic procedures? For. *Eur J Nucl Med Mol Imaging* 2004; 31(1):129-131.
22. Floriani I, Ciceri M, Torri V, Tinazzi A, Jahn H, Noseda A. Clinical profile of ioversol. A metaanalysis of 57 randomized, double-blind clinical trials. *Invest Radiol* 1996; 31(8):479-491.
23. Haynes RB, Wilczynski NL. Optimal search strategies for retrieving scientifically strong studies of diagnosis from Medline: analytical survey. *BMJ* 2004; 328(7447):1040.
24. Wilczynski NL, Haynes RB, Lavis JN, Ramkissoonsingh R, Arnold-Oatley AE. Optimal search strategies for detecting health services research studies in MEDLINE. *CMAJ* 2004; 171(10):1179-1185.

25. Wilczynski NL, Haynes RB. EMBASE search strategies for identifying methodologically sound diagnostic studies for use by clinicians and researchers. *BMC Med* 2005; 3(1):7.
26. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994; 309(6964):1286-1291.
27. Mijnhout GS, Riphagen II, Hoekstra OS. Update of the FDG PET search strategy. *Nucl Med Commun* 2004; 25(12):1187-1189.
28. Valk PE. Do we need randomised trials to evaluate diagnostic procedures? Against. *Eur J Nucl Med Mol Imaging* 2004; 31(1):132-135.
29. Bryan S, Weatherburn G, Bungay H et al. The cost-effectiveness of magnetic resonance imaging for investigation of the knee joint. *Health Technol Assess* 2001; 5(27):1-95.
30. Jarvik JG, Hollingworth W, Martin B et al. Rapid magnetic resonance imaging vs radiographs for patients with low back pain: a randomized controlled trial. *JAMA* 2003; 289(21):2810-2818.
31. Blanchard TK, Bearcroft PW, Maibaum A, Hazelman BL, Sharma S, Dixon AK. Magnetic resonance imaging or arthrography for shoulder problems: a randomised study. *Eur J Radiol* 1999; 30(1):5-10.
32. Blanchard TK, Bearcroft PW, Constant CR, Griffin DR, Dixon AK. Diagnostic and therapeutic impact of MRI and arthrography in the investigation of full-thickness rotator cuff tears. *Eur Radiol* 1999; 9(4):638-642.
33. Nikken JJ, Oei EH, Ginai AZ et al. Acute peripheral joint injury: cost and effectiveness of low-field-strength MR imaging – results of randomized controlled trial. *Radiology* 2005; 236(3):958-967.
34. Ouwendijk R, de Vries M, Pattynama PM et al. Imaging peripheral arterial disease: a randomized controlled trial comparing contrast-enhanced MR angiography and multi-detector row CT angiography. *Radiology* 2005; 236(3):1094-1103.
35. Herder GJ, Kramer H, Hoekstra OS et al. Traditional Versus Up-Front [18F] Fluorodeoxyglucose-Positron Emission Tomography Staging of Non-Small-Cell Lung Cancer: A Dutch Cooperative Randomized Study. *J Clin Oncol* 2006; .

CHAPTER 4

Design paper

Toward Less Futile Surgery in Non-Small Cell Lung Cancer? A Randomized Clinical Trial to Evaluate the Cost-Effectiveness of Positron Emission Tomography

Harm van Tinteren MSc [1], Otto S. Hoekstra MD [2,4], Egbert F. Smit MD [3], Paul Verboom MSc [5], Maarten Boers MD [4] and on behalf of the PLUS Study Group *

Comprehensive Cancer Center Amsterdam [1], Amsterdam, The Netherlands.
Departments of Nuclear Medicine [2], Pulmonology [3], and Clinical Epidemiology and Biostatistics [4], University Hospital Vrije Universiteit, Amsterdam, The Netherlands.
Institute for Medical Technology Assessment [5], Erasmus University, Rotterdam, The Netherlands.

Control Clin Trials 2001; 22: 89-98

Abstract

Non-small cell lung cancer can be cured if the patient is medically operable and the tumor resectable. Current diagnostic strategies are aimed to detect tumor deposits that preclude resection with curative intent. However, these strategies are rather inefficient, resulting in a large number of futile invasive procedures. In the early 1990s positron emission tomography. PET; showed promising results at its introduction in the clinic, especially in oncology. A large number of accuracy studies have reported that PET is superior to conventional imaging. However, whether PET ultimately improves patient outcome should ideally be assessed by means of a randomized controlled trial. No such design has been applied to evaluate PET in oncology so far. The PLUS study was designed to compare the current strategy of conventional methods with a strategy where PET was added after completion of noninvasive techniques. Patients considered operable by the physician at this point were then randomly assigned to PET and further consequences or to standard procedures of mediastinoscopy or thoracotomy. Primary outcome events were futile thoracotomies. The trial randomized 188 patients from nine hospitals in 1 year. Patient enrollment has been stopped and data collection is in progress.

Introduction

A patient with non-small cell lung cancer. NSCLC; can be cured only if the patient is medically operable and the tumor resectable. The clinical diagnostic workup of such patients aims to establish a diagnosis of NSCLC. the primary tumor; preoperatively and to identify lymphatic or hematogeneous metastases to restrict curative surgery to patients who potentially benefit from this treatment. For this purpose, a battery of diagnostic tests examining the various target tissues [lymph nodes with chest computed tomography. CT; scan, bone with plain X-rays, skeletal scintigraphy, brain with CT/magnetic resonance imaging. MRI;; liver and adrenals with CT or ultrasound] are at the clinician's disposal. Various guidelines have been formulated to optimize these efforts.[1-4] In daily clinical practice, however, variation is considerable.[5-8] In some hospitals, patients without enlarged mediastinal lymph nodes. MLN; at CT scanning proceed directly to thoracotomy. In other hospitals, mediastinoscopy is part of the standard workup. Mediastinoscopy is always performed in patients with enlarged N2 MLN. Sensitivity and specificity of CT in detecting mediastinal nodes are, respectively, 52% and 69% so that patients with negative CT who proceed to thoracotomy may in fact have N2 MLN disease[9]. Mediastinoscopy itself may also be inaccurate. Thus, the prevailing diagnostic strategy leads to futile thoracotomies in up to 50% of patients.[8,10,11]

In the early 1990s positron emission tomography. PET; emerged as a promising diagnostic imaging tool in nuclear medicine. In the oncological setting, the combination of whole-body PET with the tumor-seeking radioactive tracer fluorodeoxyglucose. ^{18}FDG; allows for noninvasive visualization and quantification of tumor deposits. Since PET provides spatial and metabolic information, it can assume a complementary role to currently available morphological imaging, or even surpass these techniques. A large number of published accuracy studies have reported, without exception, that PET is superior to the CT scan in assessing the nodal, i.e., mediastinal tumor status.[12] However, accuracy studies are not only highly susceptible to bias, but also fail to cope with the full complexities of how an imaging technique contributes to the diagnostic and therapeutic process.[13-17] By definition, the design requires that the technique be assessed out of the clinical context.[18] More accurate staging may lead to more appropriate treatment planning, but prospective evidence for that is largely anecdotal.[19] Such evidence is subsequently used in decision analyses, where risks, benefits, and costs of various strategies are modeled mathematically.[20] Several of these studies suggest that PET is cost-effective in NSCLC staging.[21,22] Given these limitations, ideally a randomized controlled trial. RCT; should assess whether PET provides information that ultimately improves patient outcome.[18,20,23]

Where patients are selected for curative surgery, the achievable health gain consists of a reduction of the number of futile surgical procedures. The PLUS study. PET in LUng cancer Staging; was designed to compare the current strategy of conventional methods with a strategy where PET was added to the noninvasive techniques with respect to these outcomes. Since the introduction of new health technology for specific clinical indications requires not only justification in terms of effectiveness but also in terms of costs, data on direct medical costs were collected concurrently.

Methodology

Population

Eight community hospitals and one academic hospital recruited patients for the trial. Patients with suspected or proven NSCLC, considered medically operable and potentially resectable by the local pulmonary physician on the basis of clinical staging procedures. i.e., clinical stage I-III;, but prior to surgical staging, were invited to participate. see Figure 1;. The diagnostic routine in each hospital was respected. Inclusion and exclusion criteria were minimal to guarantee fast accrual and generalizability. No particular adverse events were expected. Patients had to be older than 18 years of age and comply with written informed consent according to local medical ethical committee regulations.

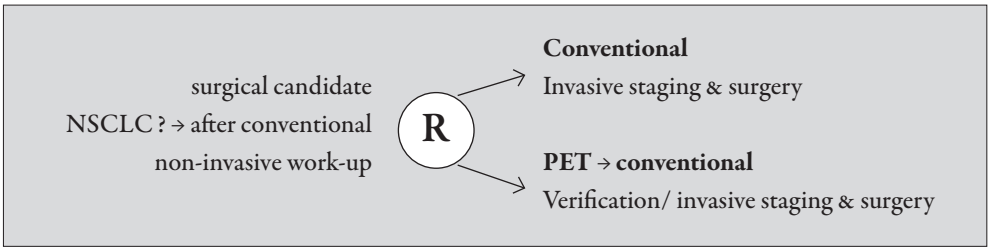


Figure 1. Trial design of PLUS study. Note that patients had to be older than 18 years and comply with written informed consent. All procedures other than PET, including therapy and follow-up, were performed by the referring institute (nine hospitals) according to prevailing local standards

In exactly one calendar year 188 patients were randomized. The results are presented in Chapter 5.

Diagnostic outcome	***Suggested implication***
CT normal and PET normal:	*Exploratory surgery with curative intent*
CT abnormal and PET normal:	
– Central tumor and adjacent lymphadenopathy	*Mediastinoscopy*
– All other cases	*Surgery with curative intent*
CT normal and PET abnormal:	*PET guided biopsy (usually mediastinoscopy)*
CT abnormal and PET abnormal:	*PET guided biopsy (usually mediastinoscopy)*

Table 1. Suggested Evaluation of the Mediastinum

Randomization

To randomize a patient, the pulmonary physicians called the central data center. After verifying eligibility criteria, patients were randomized by computer according to a permuted block design stratified for local institute. Patients were allocated either to PET scanning or to further invasive diagnostic or treatment-related procedures as dictated by local routine.

PET Procedures

PET scans were performed at the PET center of the academic hospital Vrije Universiteit in Amsterdam. The participating hospitals were located within 50 kilometers of this center. For the purpose of the study, the PET center guaranteed a PET scan within 2 days after randomization. The chest CT san was available for reading.[24] Patients had to fast for 6 hours prior to scanning. Serum glucose was measured in a blood sample drawn from an intravenous cannula that also served to inject the radioactive tracer. The PET scanner was a Siemens ECAT EXACT HR+. The 1-hour whole-body acquisition started 1 hour after injection of 370 MBq ^{18}FDG and the scanned trajectory included skull to midfemur. PET images. non-attenuation-corrected emission scans; and their interpretation were communicated to the referring physician by telephone and confirmed in writing.

PET Evaluation and Implication

Two readers of a group of three assessed all PET-scans; if necessary, consensus was reached with a third reader. Scoring of pulmonary and mediastinal lesions was visually related to the mediastinal background activity, using a three-point scale: 0 = activity less than background, 1 = activity equal to background, 2 = activity clearly above background. Transmission-emission overlays served to define the nodal status of the focal activity.[25] The first reading was blinded for CT. Focal activity outside the mediastinal area was reported on a five-point scale as normal/definitely benign, probably benign, equivocal, probably malignant, or malignant. The report of the PET scan contained information on the nature of the primary lesion, on lymph node status and distant metastases; it concluded with an assessment of TNM stage and advice for further workup in which the results of previous CT scanning were incorporated. see Table 1;. The following guideline was proposed: specimen verification of clinically relevant distant or nodal FDG foci, targeting lesion. s; with decisive impact for patient management that were the most easily accessible. Often this would imply verification of distant sites. If unconfirmed, the patient received the benefit of the doubt, and the PET findings were ignored. The attending clinicians were free to use the information and to act upon it. For patients without evidence of hematogeneous or extrathoracic lymphatic metastases the general approach described in Table 1 was proposed for evaluation of the mediastinum.[21,24]

Follow-up

All diagnostic procedures other than PET, the actual therapy as well as the clinical follow-up, took place in the referring hospital according to local standards, which consisted of regular visits. every 2-3 months; in the outpatient clinic. Follow-up information was collected 6 and 12 months after randomization. Trained data managers visited each hospital regularly and collected all data.

Outcome Events

The primary endpoint was the difference in the number of futile thoracotomies between the conventional arm and the PET-including arm. Thoracotomy was classified as futile in the case of benign lesions, pathologically proven mediastinal lymph node involvement. stage IIIA-N2;, stage IIIB, explorative thoracotomy for any other reason, or recurrent disease or death of any cause within 1 year after randomization. Thoracotomy was not considered futile in case of minimal N2-disease. i.e., intranodal involvement in a single lymph node at mediastinal dissection;. Secondary endpoints were morbidity, complications, and cost and duration of diagnostic and therapeutic processes.

Statistics

Before the trial was designed, current clinical practice was assessed in a retrospective analysis of staging procedures and their results in all patients evaluated for proven or suspected NSCLC in the two largest participating centers. Vrije Universiteit and the Medical Center Alkmaar, a community hospital.[8] Based on these results and the international PET literature, it was estimated that the use of PET imaging might reduce the number of futile thoracotomies from 45% to 20%, which corresponds to an absolute reduction of 25%. This reduction is based on the more conservative number of reduction of both hospitals. To reliably detect a reduction of 25%. power of 90%, alpha of 0.05 two-sided;, 80 evaluable patients in each arm were required. However, the potential number of patients expected from the centers in combination with the nature of the experimental technique in terms of burden to patients or physicians would allow a larger number of patients to be accrued in the trial. Therefore, the trial was open for an accrual period of 1 year, with a minimum of 160 patients. The number of futile surgical procedures in both arms will be simply expressed as a binomial value and can be tested using a Chi-square test. Logistic regression will allow the inclusion of certain potentially confounding variables, such as age of the patient and clinical stage at presentation, into the analyses. Using the definition for failures as described earlier. at surgery and 12-month follow-up;, sensitivity, specificity, accuracy, and predictive values will be calculated. Because of the expected low risk of PET imaging and the relative short accrual time, no interim statistical analysis was planned.

Cost Analysis

Data on direct medical costs of diagnosing patients. including thoracotomy; and costs due to related complications are concurrently collected for both randomization arms. Total costs, based on volume multiplied by full costs, will be calculated and compared. If PET proves to be clinically effective, sensitivity analyses based on accuracy measures provided by this study, together with information from the literature on various workup procedures in NSCLC, will be performed.

Logistics

Before investigators were allowed to randomize patients they were invited to refer at least three patients to the PET center to get acquainted with the routing and the implications of the PET scanning procedure. Meanwhile, approval of the medical ethical committee of the institute was requested. During the trial, meetings were organized every 2 months for all participants. pulmonary physicians, surgeons, radiologists; to discuss experiences and the progress of the trial. After each meeting a newsletter was sent to all study collaborators.

Discussion

The introduction of new health technology requires justification of cost-effectiveness for specific clinical indications. In the case of a diagnostic imaging technique, effectiveness means that it not only provides more accurate data than existing modalities, but also improves patient management and ultimately has a favorable impact on health status at reasonable costs. In the development of new therapeutic agents, acceptance and reimbursement can only be achieved by at least two randomized studies showing their benefit. This is in sharp contrast with the evaluation of diagnostic tests, where RCTs are extremely rare. Here, acceptance of technology seems to grow along a steadily increasing number of published accuracy studies for a certain indication. As a consequence, new and expensive but not optimally effective techniques often diffuse into clinical practice, potentially slowing down innovations in other areas.[26,27] Apart from the fact that accuracy studies are often of poor quality, their products, i.e., sensitivity and specificity estimates, are only indicative and of little relevance to the actual effectiveness of a test.[13-17,28,29] A useful test should also have diagnostic and therapeutic impact: the test should add information relevant to what was already known, and the surplus of information should contribute to therapy decisions.[14,30] Finally, even if all of these levels can be evaluated positively, the question whether the use of the technique actually contributes to the health of a patient is still unanswered. Our definition of futile thoracotomies as operationalization of health outcome of PET scanning in potentially resectable NSCLC is based on consensus of prevailing surgical management of NSCLC in the Netherlands and is supported by international guidelines.[1-3] Noncurative surgery unnecessarily increases the burden of disease and risk in many patients. In addition to surgery-related risks, life expectancy may improve if patients with locally advanced NSCLC have the opportunity to receive preoperative or neoadjuvant strategies including chemotherapy or chemoradiation.[31] We discussed including quality of life as an outcome in the trial. However, this important outcome is probably influenced by many factors outside the scope of the study. For example, it is not unlikely that quality of life would improve in the immediate aftermath of futile surgery, at least until it became obvious that the procedure failed to cure the disease. And even on failure the patient might continue to believe "that at least everything possible was done." In contrast, a patient spared futile surgery by PET would immediately experience the loss of quality associated with the verdict that cure was no longer possible or highly unlikely. Thus we felt quality of life was not useful as an outcome measure to test the efficacy of PET.

We suggest our design, i.e., a trial that randomly allocates patients to either conventional diagnostic workup or similar workup supplemented with the new technique, may be optimal to answer this question. The major benefit of randomization in a diagnostic process is that at any moment all available information on a particular patient can be used and that the randomized groups are similar with respect to all known and unknown factors, thus permitting an unbiased comparison. It is difficult to reason why these designs are applied so rarely in the evaluation of diagnostic techniques. Obviously, a complexity of forces are involved in the assessment and diffusion of new technologies.[32] "Publishing early and often by academics fortunate enough to gain early access to new technology has been rewarded with accelerated promotion and tenure, recognition within the diagnostic community, and often, favorable financial arrangements with manufactures eager to align themselves with well-known scientists".[33] Because of this uncritically low level of acceptance of new technologies, the time window in which investigators consider it ethical to randomize patients is extremely short.[34] Medical ethical committees, policy makers, and patient movements might also put pressure on this window of opportunity.[31,35] For the present study, patient enrollment was finished within 1 year. Data from the Netherlands Cancer Registry[36] indicate that the present study included approximately 65% of all eligible patients diagnosed in the nine participating centers. In the Netherlands, PET is not yet commonly available for routine diagnostic investigations. PET capacity is still limited to three scanners per 15 million inhabitants. Because of this shortage of availability, the ethics of this trial are justified, as was the case in the pioneering MRC trial of streptomycin for tuberculoses.[37] Hopefully, this trial is still timely and may help the clinical community in defining guidelines for cost-effective use for optimal application of this technique.

References

1. N. Van Zandwijk, Consensus bijeenkomst diagnostiek longcarcinoom. *Ned Tijdschr v Geneesk*. 1991; 135: pp. 1915-1919.
2. P. Goldstraw, P. Rocmans, D. Ball *et al.*, Pretreatment minimal staging for non-small cell lung cancer: An updated consensus report. *Lung Cancer*. 1994; 11 Suppl 3: pp. S1-S4.
3. National Comprehensive Cancer Network, Non-small-cell lung cancer practice guidelines. *Oncology*. 1996; 10 Suppl: pp. 81-111.
4. Anonymous, Pretreatment evaluation of non-small-cell lung cancer. *Am J Respit Crit Care Med*. 1997; 156: pp. 320-322.
5. G.H. Guyatt, D.J. Cook, L.E. Griffith *et al.*, Surgeons' assessment of symptoms suggesting extrathoracic metastases in patients with lung cancer. Canadian Lung Oncology Group. *Ann Thorac Surg*. 1999; 68: pp. 309-315.
6. W.R. Webb and J.A. Golden, Imaging strategies in the staging of lung cancer. *Clinics in Chest Medicine*. 1991; 12: pp. 133-150.
7. R.J.A.U. Fergusson, A.A.U. Gregor, R.A.U. Dodds *et al.*, Management of lung cancer in South East Scotland. *Thorax*. 1996; 51: pp. 569-574.
8. G.J.M. Herder, C.D. Colder, I. van Mansom *et al.*, Staging non-small cell lung cancer patients in two Dutch hospitals. *Eur J Resp Dis*. 1999; 14: p. 446.
9. W.R. Webb, C. Gatsonis, E.A. Zerhouni *et al.*, CT and MR imaging in staging non-small cell bronchogenic carcinoma: Report of the Radiologic Diagnostic Oncology Group. *Radiology*. 1991; 178: pp. 705-731.
10. P. Goldstraw, The practice of cardiothoracic surgeons in the perioperative staging of non-small cell lung cancer. *Thorax*. 1992; 47: pp. 1-2.
11. The Canadian Lung Oncology Group, Investigation for mediastinal disease in patients with apparently operable lung cancer. *Ann Thorac Surg*. 1995; 60: pp. 1382-1389.
12. B.A. Dwamena, S.S. Sonnad, J.O. Angobaldo and R.L. Wahl, Metastases from non-small cell lung cancer: Mediastinal staging in the 1990s – Meta-analytc comparison of PET and CT. *Radiology*. 1999; 213: pp. 530-536.
13. C.B. Begg, Biases in the assessment of diagnostic tests. *Stat Med*. 1987; 6: pp. 411-423.
14. J. Cormack, C.A. Evill, S.P. Langlois, M.R. Sage and A.M. Tordoff, Evaluating the clinical efficacy of diagnostic imaging procedures. *Eur J Radiol*. 1992; 16: pp. 1-9.
15. M.C. Reid, M.S. Lachs and A.R. Feinstein, Use of methodological standards in diagnostic research. Getting better but still not good. *JAMA*. 1995; 274: pp. 645-651.
16. S. Kelly, E. Berry, P. Roderick *et al.*, The identification of bias in studies of the diagnostic performance of imaging modalities. *Br J Radiol*. 1997; 70: pp. 1028-1035.
17. Adams E, Flynn K. Positron Emission Tomography. Descriptive Analysis of Experience with PET in VA. A Systematic Review Update of FDG-PET as a Diagnostic Test in Cancer and Alzheimer's Disease. Report 10. December 1998;. MDRC, VA Medical Center Boston.
18. L.S. Freedman, Evaluating and comparing imaging techniques: A review and classification of study designs. *BJR*. 1987; 60: pp. 1071-1081.
19. P. Lewis, S. Griffin, P. Marsden *et al.*, Whole-body 18F-fluorodeoxyglucose positron emission tomography in pre-operative evaluation of lung cancer. *Lancet*. 1994; 344: pp. 1265-1266.
20. M.G.M. Hunink, Outcomes research and cost-effectiveness analysis in radiology. *Eur Radiol*. 1996; 6: pp. 615-620.
21. S.S. Gambir, C.G. Hoh, M.E. Phelps *et al.*, Decision tree sensitivity analysis for cost-effectiveness of FDG-PET in the staging and management of non-small cell lung carcinoma. *J Nucl Med*. 1996; 37: pp. 1428-1436.
22. W.J. Scott, J. Shepherd and S.S. Gambhir, Cost-effectiveness of FDG-PET for staging non-small cell lung cancer. *Ann Thoracic Surg*. 1998; 66: pp. 1428-1436.

23. N.S. Weiss, Diagnostic and screening tests: Measuring their role in improving the outcome of illness. In: Clinical Epidemiology: The Study of the Outcome of Illness. *Monographs in Epidemiology and Biostatistics*. 1996; 27: pp. 27-46.
24. J.F. Vansteenkiste, S.G. Stroobants, P.J. Dupont *et al.*, FDG-PET scan in potentially operable non-small cell lung cancer: Do anatometabolic PET-CT fusion images improve the localisation of regional lymph node metastases? The Leuven Lung Cancer Group. *Eur J Nucl Med*. 1998; 25: pp. 1495-1501.
25. T. Naruke, K. Suematsu and S. Ishikawa, Lymph node mapping and curability at various levels of metastasis in resected lung cancer. *J Thorac Cardiovasc Surg*. 1978; 76: pp. 832-839.
26. L. Dalla-Palma, A.K. Dixon, I. Durand-Zaleski *et al.*, An overview of cost-effective radiology. *Eur Radiol*. 1997; 7: pp. 147-150.
27. A.K. Dixons, Evidence-based diagnostic radiology. *Lancet*. 1997; 350: pp. 509-512.
28. D.W. Gelfland and D.J. Ott, Methodologic issues in comparing imaging methods. *AJR*. 1985; 144: pp. 1117-1121.
29. L.S. Cooper, T.C. Chalmers, M. McCally, J. Berrier and H.S. Sacks, The poor quality of early evaluations of magnetic resonance imaging. *JAMA*. 1988; 259: pp. 3277-3280.
30. R. Mackenzie and A.K. Dixon, Measuring the effects of imaging: An evaluative framework. *Clin Radiol*. 1995; 50: pp. 513-518.
31. P.C. Hoffman, A.M. Mauer and E.E. Vokes, Lung Cancer. *Lancet*. 2000; 355: pp. 479-485.
32. R. Rosen and J. Gabbay, Linking health technology assessment to practice. *BMJ*. 1999; 319: pp. 1292-1294.
33. B.J. Hillman, Outcomes research and cost-effectiveness analysis for diagnostic imaging. *Radiology*. 1994; 193: pp. 307-310.
34. G. Mowatt, D.J. Bower, J.A. Brebner *et al.*, When and how to assess fast-changing technologies: A comparative study of medical applications of four generic technologies. *Health Technology Assessment*. 1997; 1: p. 14.
35. Anonymous. Imaging for hope. Woman's voice for positron emission tomography. Supported by Jennifer Jones Simon Foundation. UCLA School for Medicine. Institute for Clinical PET, U.S. Department of Energy. 1999.
36. O. Visser, J.W.W. Coebergh, L.J. Schouten and J.A.A.M. van Dijck, Editors, *Incidence of Cancer in the Netherlands 1996*, VICK, Utrecht. 2000;.
37. Medical Research Council, Streptomycin treatment of pulmonary tuberculosis. *BMJ*. 1948; ii: pp. 769-782.

* Participating pulmonologists of the PLUS study are: J.H.A.M. van den Bergh, Medical Centre Alkmaar; R.A.L.M. Stallaert, Westfries Gasthuis, Hoorn; A.J.M. Schreurs, Onze Lieve Vrouwe Gasthuis, Amsterdam; J.P. Teengs, Kennemer Gasthuis, Haarlem; J. Berkovits, Ziekenhuis Amstelveen; W.F. Strankinga, BovenIJ Ziekenhuis, Amsterdam; G. Visschers, Slotervaart Ziekenhuis, Amsterdam; and C. Jie, Lucas/Andreas Ziekenhuis, Amsterdam.

CHAPTER

5

Effectiveness of positron emission tomography in the preoperative assessment of patients with suspected non-small-cell lung cancer: the PLUS multicentre randomised trial

Harm van Tinteren MSc[1], Otto S Hoekstra MD[2, 4], Egbert F Smit MD[3], Jan HAM van den Bergh MD[6], Ad JM Schreurs MD[7], Roland ALM Stallaert MD[8], Piet CM van Velthoven MD[5], Emile FI Comans MD[2], Fred W Diepenhorst[1], Paul Verboom MSc[8], Johan C van Mourik MD[5], Prof Pieter E Postmus MD[3], Prof Maarten Boers MD[4], Prof Gerrit JJ Teule MD[2] and the PLUS study group

Members of the PLUS study group are listed at the end of the report

Comprehensive Cancer Centre Amsterdam[1], Amsterdam, Netherlands; Department of Nuclear Medicine[2], Pulmonology[3], Clinical Epidemiology and Biostatistics[4], Surgery[5], VU University Medical Centre, Amsterdam; Department of Pulmonology Medical Centre Alkmaar[5], Alkmaar; Department of Pulmonology Onze Lieve Vrouwe Gasthuis[6], Amsterdam; Department of Pulmonology, Westfries Gasthuis Hoorn[7], Hoorn; Institute for Medical Technology Assessment[8], Erasmus University, Rotterdam

The Lancet 2002; 359:1388-1392

Summary

Background

Up to 50% of curative surgery for suspected non-small-cell lung cancer is unsuccessful. Accuracy of positron emission tomography (PET) with 18-fluorodeoxyglucose (18FDG) is thought to be better than conventional staging for diagnosis of this malignancy. Up to now however, there has been no evidence that PET leads to improved management of patients in routine clinical practice. We did a randomised controlled trial in patients with suspected non-small-cell lung cancer, who were scheduled for surgery after conventional workup, to test whether PET with 18FDG reduces number of futile thoracotomies.

Methods

Before surgery (mediastinoscopy or thoracotomy), 188 patients from nine hospitals were randomly assigned to either conventional workup (CWU) or conventional workup and PET (CWU+PET). Patients were followed up for 1 year. Thoracotomy was regarded as futile if the patient had benign disease, explorative thoracotomy, pathological stage IIIA-N2/IIIB, or postoperative relapse or death within 12 months of randomisation. The primary outcome measure was futile thoracotomy. Analysis was by intention to treat.

Findings

96 patients were randomly assigned CWU and 92 CWU+PET. Two patients in the CWU+PET group did not undergo PET. 18 patients in the CWU group and 32 in the CWU+PET group did not have thoracotomy. In the CWU group, 39 (41%) patients had a futile thoracotomy, compared with 19 (21%) in the CWU+PET group (relative reduction 51%, 95% Cl 32-80%; p = 0·003).

Interpretation

Addition of PET to conventional workup prevented unnecessary surgery in one out of five patients with suspected non-small-cell lung cancer.

Introduction

Accurate staging of patients with a pulmonary lesion suspected of being non-small-cell lung cancer is needed to restrict surgical or multimodality treatment to those who will potentially benefit from these treatments. Several imaging techniques and invasive tests are available to the clinician to detect mediastinal lymph-node involvement, distant metastases, or both. International guidelines to make the most of this process[1-4] have been formulated, but routine clinical practice remains variable.[5-8] Despite current diagnostic workup, early local and distant relapses are frequent, and surgery is done for preoperatively suspicious lesions that can prove to be benign. Therefore, surgery can be regarded as futile in up to 50% of patients with presumably resectable non-small-cell lung cancer.[8-10]

Positron emission tomography (PET) with the tracer 18-fluorodeoxyglucose (18FDG) has emerged in the past decade as a promising oncological imaging tool. Results of several accuracy studies have suggested that 18FDG-PET is better at assessment of suspicious lung lesions and nodal or extra-thoracic tumour status in non-small-cell lung cancer than conventional workup.[11-14] Accuracy studies are, however, not designed to show added value of diagnostic tests. Like phase II studies for development of treatments, they are subject to bias, which make generalisation of results to predict an effect in routine practice difficult.[15] In general, these drawbacks lead to overestimation of worth.[16] As a result, whether and to what extent patients will benefit from use of PET in a routine clinical setting cannot be directly inferred from existing published work.[17,18]

As in assessment of new treatments, new diagnostic technologies need to be compared with current strategies with respect to relevant clinical outcomes.[19-21] Workers on major health-technology assessment reports[20,22] concluded that improvement of diagnostic accuracy by PET was difficult to quantify because of variable quality of studies, and that direct evidence on the effect of PET in improvement of patients' outcomes was still lacking.

The PLUS (PET in LUng cancer Staging) study was designed to work with routine clinical workup of patients with suspected non-small-cell lung cancer. We compared the current strategy of conventional diagnostic methods with a strategy in which PET was added to non-invasive diagnostic techniques. The primary outcome measure was number of futile thoracotomies.

Patients and methods

Patients

We invited patients with suspected or proven non-small-cell lung cancer that was judged to be medically operable and potentially resectable by the referring clinician on the basis of clinical staging—but not surgical staging—to participate in the study. Eligible patients had to be older than age 18 years. All patients gave written informed consent in accordance with requirements set by local medical ethics committees. Eight community and one university hospital recruited patients for the study.

Procedures

We randomly assigned eligible patients either PET followed by further invasive diagnostic and therapeutic procedures (CWU+PET) or invasive diagnostic and therapeutic procedures alone (CWU). These procedures are governed by local routine, which is based on current guidelines.[1-4] Randomisation was done centrally by computer, by a permuted block design, stratified by institute.

All procedures other than PET, including treatment and follow-up, were done in the referring hospitals according to local standards. Follow-up consisted of regular visits (at least every 2–3 months) in the outpatient clinic. Trained data managers obtained all information up to 12 months after randomisation.

We did PET scans with a Siemens ECAT EXACT HR+scanner (Siemens/CTI, Knoxville, TN, USA) at the Vrije Universiteit, Amsterdam. We asked patients to fast for 6 h before the scan. A 1-h whole-body acquisition started 60 min after injection of 370 MBq 18FDG. We imaged the mid-femur-skull trajectory with emission scans for 5 min except for the chest, which was imaged with two 10-min scans followed by 5 min of transmission scanning. Images were reconstructed by filtered back-projection (Hanning 0·5; resolution after reconstruction 7 mm full-width at half-maximum) without attenuation correction. Results were communicated to the referring clinician by phone and confirmed in writing, with a hard copy of the PET scan included.

Two readers from a group of three assessed all PET scans; if necessary, consensus was reached with the third reader. Pulmonary and mediastinal lesions were visually associated with mediastinal background activity.[23] Chest computed tomography (CT) scans and transmission-emission overlays defined anatomical correlates for PET abnormalities and localised mediastinal foci.[24]

The PET report included information on nature of the primary lesion, mediastinal lymph-node involvement, and distant metastases, and concluded with an assessment of tumour-node metastasis (TNM) stage according to CT and PET and a suggestion for further workup. Referring clinicians used this information accordingly. However, potential distant or nodal metastases, which might have a major effect on patients' management, were confirmed by other means. Unconfirmed PET findings were ignored.

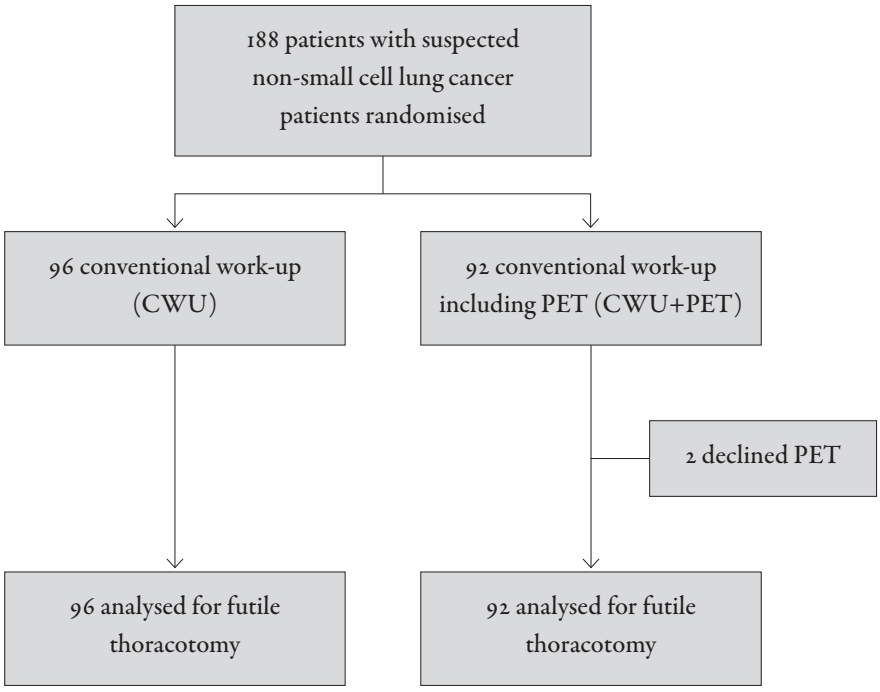
The primary outcome measure was number of futile thoracotomies. We classified thoracotomy as futile for the following reasons: benign lung lesion; pathologically proven mediastinal lymph-node involvement (stage IIIA–N_2)[25] other than minimal N_2-disease—ie, intranodal involvement in a

single lymph node established at mediastinal dissection;[26] stage IIIB disease; explorative thoracotomy for any other reason; or recurrent disease or death from any cause within 1 year after randomisation.

Statistical analysis

Before the trial was designed, we assessed current clinical practice in a retrospective analysis of staging procedures in the two largest participating centres.[8] On the basis of these results and of published work about PET, we estimated that addition of PET to conventional workup might reduce the number of futile thoracotomies from 45% to 20%. Our sample-size calculation (power of 90%, of 0·05, two-sided) suggested that a total of 160 patients was required.[27] Number of futile thoracotomies with respect to total number of patients randomised in each group was tested by $\chi 2$ test (intention-to-treat analysis). The same test compared nodal yield between both groups. In addition, we did tests that were based on exact inferences and that were stratified by institute. None of these approaches, however, led to important changes in results. Hence, for reasons of arithmetic simplicity, we report unstratified $\chi 2$ tests and their CIs. Because of the low risk of PET imaging and the short accrual time, no interim statistical analysis was planned.

FIGURE. Trial profile



Results

Baseline characteristics

Between January, 1998, and January, 1999, we enrolled 188 patients from nine hospitals (between five and 50 patients per hospital) – 96 in the CWU group and 92 in the CWU+PET group (figure). 70% of patients in each group had clinical stage I/II disease (table 1). Pre-randomisation workup was closely similar for both groups, and included at least a chest CT scan, which usually included

	CWU (n = 96)	CWU+PET (n = 92)
Characteristic		
Age (years, mean [SD])	65 (10)	66 (10)
Sex		
Men	75 (78%)	69 (75%)
Women	21 (22%)	23 (25%)
Karnofsky index		
70-80	6 (6%)	10 (11%)
90-100	90 (94%)	82 (89%)
Weight loss >5%	15 (16%)	14 (15%)
Clinical stage		
I	63 (66%)	58 (63%)
II	5 (5%)	6 (7%)
IIIA	22 (23%)	23 (25%)
IIIB	6 (6%)	4 (4%)
IV*		1 (1%)
Comorbidity		
Vascular, diabetes mellitus	31 (32%)	30 (33%)
Chronic obstructive pulmonary disease	30 (31%)	23 (25%)
Previous malignancies	13 (14%)	17 (19%)
Definite diagnosis of non-small-cell lung cancer	46 (48%)	48 (52%)
Pre-randomisation imaging tests (CT of the thorax excluded)	56 (58%)	54 (59%)
Bone scan	26 (27%)	25 (27%)
CT/US of the abdomen	46 (48%)	42 (46%)
CT/MRI of the brain	3 (3%)	5 (5%)
CT/MRI of other area	1 (1%)	5 (5%)
Radiograph of other area	7 (7%)	10 (11%)

Data are number of patients (%) unless otherwise stated. CT = computed tomography; US = ultrasound; MRI = magnetic resonance imaging.
* Solitary brain metastasis on CT.

Table 1. Baseline characteristics

the liver and adrenals (89%). All CT scans were done with at least third-generation scanners: spiral modality in 24%, with intravenous contrast in 55%, both equally distributed between each group. In both groups, 58% of patients underwent at least one additional test to identify metastatic disease (table 1). Two patients allocated to the CWU+PET group declined PET (figure). The other 90 patients underwent PET a median of 3 days after randomisation (range 1-13).

Primary outcome

A significantly greater proportion of patients underwent futile thoracotomy in the CWU group than in the CWU+PET group (relative reduction 51%, 95% CI 32–80, p = 0·003; table 2). The absolute difference of 20% can be interpreted as five patients (95% CI 3–14) who needed PET to avoid one futile thoracotomy. Apparently justified surgery was done in a closely similar proportion of patients in each group (table 2). Futile surgery happened irrespective of clinical stage: 46% (31/68) of patients with clinical stage I/II disease in the CWU group compared with 25% (16/64) of those in the CWU+PET group, and 29% (8/28) of those in the CWU group with clinical stage III disease compared with 11% (3/27) in the CWU+PET group, had futile surgery. Assessment of resectability by CT and PET was discordant in a third of cases, and PET was correct in two-thirds. In 37% of patients undergoing PET, CT was incorrect with respect to the primary outcome measure. This percentage was closely similar to the proportion of futile thoracotomies in the CWU group.

Our protocol required confirmation of clinically decisive PET results, thus the predictive value of PET alone (ie, without any confirmation) was analysed in 86 assessable patients. Six patients could not be assessed because they either declined PET (n = 2), had intercurrent morbidity prohibiting surgery (2), refused surgery (1), or had chemotherapy rather than surgery after revision of tumour histology (1). Six patients who died after apparently curative surgery were allocated to justified thoracotomy (table 3). PET correctly suggested that surgery was justified in 81% (95% CI 68–92) of scans, versus 71% (58–84%) in which PET suggested surgery was futile. Overall accuracy before verification was 76% (67–85%). Better preoperative assessment of patients in the CWU+PET group than in the CWU group was evident in three phases: before surgery with curative intent, at surgery, and during follow-up.

Before surgery with curative intent

After randomisation, 18 patients in the CWU group did not proceed to thoracotomy for the following reasons (table 2): tumour-positive mediastinal lymph-node biopsy sample (n = 10); shrinking lesion at preoperative chest radiograph (2); adrenal metastasis diagnosed by revision of CT confirmed by fine-needle aspiration (1); small-cell lung cancer (2); cardiac disease prohibiting surgery (1); death due to local progression before surgery (1); and surgery declined (1). Thus, 78 patients underwent thoracotomy with curative intent.

63 patients (66%) underwent 67 mediastinal lymph-node staging attempts: mediastinoscopy, 62; mediastinotomy, four; video-assisted thoracoscopy, one. Of the ten patients with biopsy-proven mediastinal lymph-node involvement precluding surgery, in five, the final N-stage was diagnosed at CT-indicated mediastinal lymph-node stations; in two, CT suggested absent and hilar adenopathy;

and in three, confirmation was obtained at stations not enlarged at CT, one of which proved to have a contralateral-positive biopsy, rather than only ipsilateral involvement, as suggested by CT. 25 patients underwent thoracotomy without additional staging procedures after randomisation: 22 had cT1/2N0M0, two had cT2N2 (nodes not accessible by mediastinoscopy, and mediastinoscopy impossible after previous neck surgery, respectively), and one had cT3N0.

In the CWU+PET group, 32 patients did not proceed to thoracotomy for the following reasons (table 2): tumour-positive lymph nodes (n = 18; two after PET-guided lower cervical node biopsies); confirmed stage IV disease after PET (7; skeletal, liver, and adrenal metastasis confirmed by biopsy; cerebral metastasis by CT and magnetic resonance imaging); PET and clinical course suggesting benign disease (3; two had a normal PET scan); intercurrent morbidity prohibiting surgical treatment (2); biopsy-confirmed neuroendocrine tumour treated with chemotherapy (1); and surgery declined (1). Thus, 60 patients in the CWU+PET group underwent thoracotomy with curative intent.

67 patients (73%) had 68 mediastinal lymph-node staging attempts: mediastinoscopy, 59; mediastinotomy, five; and rigid bronchoscopy, four. Of the 18 patients with biopsy-proven mediastinal lymph-node involvement precluding surgery, pathological N-classification was established at a CT-enlarged nodal station in eight versus a PET-positive mediastinal lymph node in 16. In six patients, the final N-stage was diagnosed at sites shown by CT and PET (2N3, 4N2). In ten patients, only PET suggested the positive biopsy site (4N3, 6N2), five of which had a normal CT scan. In two patients, PET had suggested hilar adenopathy (N1) with positive biopsies of ipsilateral lower tracheobroncheal nodes (with CT read as N0 and N1, respectively). In three patients, the mediastinal lymph-node procedure had been done while awaiting confirmation of coexistent PET-suspected distant sites. Number of nodal stations sampled at mediastinoscopy was closely similar in each group.

In 13 patients without evidence of mediastinal lymph-node involvement on chest CT, PET suggested otherwise, which was confirmed in six. Four patients could not be assessed for various reasons (patient refusal, exploratory surgery), and at thoracotomy, no mediastinal lymph-node involvement was noted in the remaining three patients. False-positive PET-suspected rib metastases were recorded in two patients (fibrodysplasia, trauma). 18 patients underwent thoracotomy without additional staging procedures after randomisation and PET: 17 had clinical stage cT1/2N0/1 and one cT1N2 (nodes inaccessible by mediastinoscopy).

At surgery

Patients in the CWU and CWU+PET groups underwent thoracotomy at a median of 22 days (range 8–118) and 28 days (4–106), respectively. Number of patients proceeding to thoracotomy was significantly higher in the CWU group than in the CWU+PET group (table 2; p = 0·013). In the CWU group, 20 thoracotomies were futile versus nine in the CWU+PET group. 12 patients who had CWU were upstaged and seven had benign disease (table 2). In the CWU+PET group, six patients were upstaged and two had benign disease (table 2). In three of the patients with IIIA–N2 disease, PET suggested mediastinal lymph-node involvement, which was not confirmed by mediastinoscopy. In both groups, one open-and-close procedure was done, since residual lung capacity precluded the pneumonectomy that was necessary to achieve radical surgery.

In total, a closely similar number of different nodal stations were sampled in CWU and CWU+PET groups (mean 4·3 [SD 1·9] *vs* 4·6 [1·8], respectively). The yield of preoperative mediastinal lymph-node staging was 63% in the CWU group compared with 83% in the CWU+PET group (p = 0·16).

	CWU (n = 96)	CWU+PET (n = 92)
No thoracotomy	18 (19%)	32 (35%)
Confirmed N2/3	10	18
Confirmed distant metastases	1	7
Benign primary lesions	2	3
Other tumour	2	1
Intercurrent morbidity, refusal	3	3
Thoracotomy	78 (81%)	60 (65%)
Non-futile thoracotomy	39 (41%)	41 (44%)
Futile thoracotomy	39 (41%)	19 (21%)
Benign	7	2
Explorative thoracotomy	1	1
IIIA–N2	6	4
IIIB	6	2

Table 2. Specification of primary outcome

During follow-up

In the CWU group, 14 patients developed clinically noticeable recurrences within 12 months of randomisation after apparently curative surgery. Patients relapsed with metastases in the brain (n = 3), lymph nodes (2), bone (2), kidney (2), soft tissue (3), adrenal gland (1), and liver (1), with multiple sites in four patients. Most relapses arose beyond 180 days of randomisation. Nine of the 14 patients who relapsed died during follow-up. Another four died of surgery-related causes, without clinical evidence of relapse. One patient died of reasons not definitely due to cancer or surgery. Of all patients who underwent surgery in the CWU group, 17 had distant metastases within the year of follow-up. In all but two patients, these sites had been screened before randomisation, or there had been no indication for such screening according to 1997 guidelines (n = 5).[4]

In the CWU+PET group, four patients relapsed after apparently curative surgery, two with bone metastases, and one with metastases in the brain and skeleton after refusal of a PET scan and undergoing curative surgery (pT1N0M0). In the other patient, a pulmonary metastasis of melanoma (primary site unknown) had been resected, but disseminated involvement (skin, breast, adrenal) became apparent during follow-up. All patients with recurrent disease were alive 12 months after randomisation. Five patients died of surgery-related causes and one patient died of an unknown cause.

	Thoracotomy		No thoracotomy		Total
	Justified	Futile	Justified	Futile	
Thoracotomy indicated by PET					
Yes	33	6*	2†	0	41
No	13‡	6§	26¶	0	45

* Benign disease in two patients, advanced disease in four.
† PET suggested hilar lymph-node involvement, but mediastinoscopy was positive.
‡ PET suggested benign disease in one, advanced disease in 12 (including patients in whom PET could not exclude mediastinal lymph-node involvement adjacent to the primary tumour.
§ Thoracotomy and follow-up showed advanced disease.
¶ PET suggested benign disease in three, advanced disease in 23.

Table 3. Accuracy of PET in prediction of need for thoracotomy

Discussion

Our study showed that addition of PET to conventional workup can strikingly reduce the number of futile thoracotomies in patients with suspected potentially resectable non-small-cell lung cancer. The main effect of PET was to upstage patients (12% in the CWU group compared with 27% in the combined group). Obviating surgery in such patients improves patients' management.

Our findings are directly applicable to clinical practice. Data from the Netherlands Cancer Registry[28] suggest that our study probably included about 65% of all eligible patients diagnosed in these nine hospitals.

Our definition of futile thoracotomy as operationalisation of health outcome of PET in potentially resectable non-small-cell lung cancer is based on consensus of current surgical management of non-small-cell lung cancer in the Netherlands, and is supported by international guidelines.[1,3] Non-curative surgery unnecessarily increases burden of disease and risk. The life expectancy of patients with locally advanced non-small-cell lung cancer might improve if they receive preoperative or neoadjuvant treatments, including chemotherapy or chemoradiation.[29] If resection in any cancer patient who survived clinically disease-free for 1 year had been deemed to be justified (rather than futile, classified in this trial by perioperative IIIA-N2/IIIB stage), the overall conclusion would be much the same (39% *vs* 16% futile surgery in the CWU and CWU+PET groups, respectively; p = 0·001). Likewise, if surgery for benign disease and non-cancer-related death were excluded, there would still be a difference (29% *vs* 13% respectively; p = 0·007). Only one patient (in the CWU+PET group) was reported with minimal N2 disease at thoracotomy.

Unexpected distant metastases during follow-up were noted in tissues that had been screened or in patients who were not at risk in accordance with the guidelines.[4] Our finding of 8% confirmed distant metastases in the CWU+PET group is slightly below figures from other studies,[12,13,30,31] in which 11–14% unexpected distant metastases have been reported.

In retrospect, mediastinal lymph-node staging procedures were fully compliant with 1997 international guidelines[4] and were done in about two-thirds of patients. In the CWU group, the sensitivity of these efforts was about 60%, which is closely similar to results from other multicentre trials.[10] Although our trial was not designed to address the issue, our data suggest a high yield of invasive preoperative mediastinal lymph-node staging if guided by PET of suspected nodes. At the time when our trial was designed (in 1997),[32] the interaction between PET and mediastinoscopy was unclear. Rigorous accuracy studies[13,33] have since shown that the negative predictive value of PET for mediastinal lymph-node involvement could be sufficiently high to refrain from mediastinoscopy in non-central tumours.

Some aspects typically associated with management of patients, as opposed to diagnostic accuracy, became apparent during our trial. First, because 18FDG uptake does not always suggest malignancy, and surgery is the only chance of cure for most patients with non-small-cell lung cancer, PET findings that would preclude surgical treatment need to be verified. This requirement is supported by the estimated diagnostic accuracy of PET alone in this study (table 3). The claimed superior accuracy – eg, mediastinal lymph-node involvement versus CT[11,13] – needs to be translated into relevant outcome measures, ie, into cases upstaged by preoperative biopsy rather than at surgery with curative intent. Even if false-positive PET findings might not adversely affect the ultimate result of workup, they are likely to invoke additional tests, generating an unnecessary burden on the patient, delay, and costs.

Second, clinical decision-making takes into account the complete diagnostic profile of the patient, and not merely the result of a single test. Therefore, a patient with a clinical and radiological profile strongly suggesting malignancy underwent thoracotomy despite a PET-negative primary lesion. This decision was justified in retrospect (diagnosis of bronchioloalveolar cell cancer). Finally, in some cases, preoperative radiological follow-up already suggested benign disease and obviated surgery. PET would not have affected this decision. Such data cannot be derived from accuracy studies because of masking for the new procedure, and are difficult to model in decision analysis.[34]

In conclusion, addition of PET to standard workup in routine clinical practice improved selection of surgically curable patients with non-small-cell lung cancer.

PLUS study group participants

A Boonstra, B Venmans, A J M van Boxem, G Sutedja, R A Manoliu, R Golding (Academic Hospital Vrije Universiteit, Amsterdam); W F M Strankinga, P M Hooghiemstra, F C Crezée, P L Tolenaar (BovenIJ Ziekenhuis, Amsterdam); H B Kwa, J G van Unnik, R Hardjowijono, S S Wagenaar (Onze Lieve Vrouwe Gasthuis, Amsterdam); C Jie, M C T B Sie, R M J M Butzelaar, M N Weimann (Lucas Ziekenhuis, Amsterdam); G Visschers, P I van Spiegel, G C Collet, R P Rademakers (Slotervaart Ziekenhuis, Amsterdam); W G Boersma, M Deenstra, C S de Graaff, T Haitjema, G H Ooms, J H Pot (Medisch Centrum Alkmaar); J Prins, P M J M de Vries, E Scheijde, J Dijkstra (Westfries Gasthuis, Hoorn); J P Teengs, E Boerma (Kennemer Gasthuis, Haarlem); J Berkovits, R P A Boom, J Krekt (Ziekenhuis Amstelveen); E W M Grijseels (Institute for Medical Technology Assessment, Erasmus University, Rotterdam).

References

1. Zandwijk van N. Consensus bijeenkomst diagnostiek longcarcinoom. Ned Tijdschr Geneeskd 1991; 135: 1915-1919.
2. Goldstraw P, Rocmans P, Ball D, et al. Pretreatment minimal staging for non-small cell lung cancer: an updated consensus report. *Lung Cancer* 1994; 11 (suppl 3): S1-S4.
3. National Comprehensive Cancer Network. Non-small-cell lung cancer practice guidelines. *Oncology* 1996; 10 (suppl): 81-111.
4. Anon. Pretreatment evaluation of non-small-cell lung cancer. *Am J Respir Crit Care Med* 1997; 156: 320-322.
5. Guyatt GH, Cook DJ, Griffith LE, et al. Surgeons' assessment of symptoms suggesting extrathoracic metastases in patients with lung cancer. *Ann Thorac Surg* 1999; 68: 309-315.
6. Webb WR, Golden JA. Imaging strategies in the staging of lung cancer. *Clin Chest Med* 1991; 12: 133-150.
7. Fergusson RJ, Gregor A, Dodds R, Kerr G. Management of lung cancer in South East Scotland. *Thorax* 1996; 51: 569-574.
8. Herder GJM, Colder CD, van Mansom I, et al. Staging non-small cell lung cancer patients in two Dutch hospitals. *Eur Respir J* 1999; 14: 446.
9. Goldstraw P. The practice of cardiothoracic surgeons in the perioperative staging of non-small cell lung cancer. *Thorax* 1992; 47: 1-2.
10. The Canadian Lung Oncology Group. Investigation for mediastinal disease in patients with apparently operable lung cancer. *Ann Thorac Surg* 1995; 60: 1382-1389.
11. Dwamena BA, Sonnad SS, Angobaldo JO, Wahl RL. Metastases from non-small cell lung cancer: mediastinal staging in the 1990s—meta-analytic comparison of PET and CT. *Radiology* 1999; 213: 530-536.
12. Bury T, Dowlati A, Paulus P, et al. Whole-body 18FDG positron emission tomography in the staging of non-small cell lung cancer. *Eur Respir J* 1997; 10: 2529-2534.
13. Pieterman RM, van Putten JWG, Meuzelaar JJ, et al. Preoperative staging of non-small cell lung cancer with positron emission tomography. *N Engl J Med* 2000; 343: 254-261.
14. Lowe VJ, Fletcher JW, Gobar L, et al. Prospective evaluation of positron emission tomography in lung nodules. *J Clin Oncol* 1998; 16: 1075-1084.
15. Begg CB. Biases in the assessment of diagnostic tests. *Stat Med* 1987; 6: 411-423.
16. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999; 282: 1061-1066.
17. Spagnolo SV. The diagnostic strategy for lung cancer. *Chest* 2000; 117: 1219-1220.
18. Lewis P, Griffin S, Marsden P, et al. Whole-body 18F-fluorodeoxyglucose positron emission tomography in preoperative evaluation of lung cancer. *Lancet* 1994; 344: 1265-1266.
19. Weiss NS. Diagnostic and screening tests: measuring their role in improving the outcome of illness In: . In: Weiss, NS (Ed.), Clinical epidemiology: the study of the outcome of illness. vol 27:New York: Oxford University Press, 1996: 27-46.
20. Adams E, Flynn K. Descriptive analysis of experience with PET in VA: a systematic review update of FDG-PET as a diagnostic test in cancer and Alzheimer's disease

    http://www.va.gov/resdev/prt/petreport.htm. (accessed Dec 19, 2001).
21. Freedman LS. Evaluating and comparing imaging techniques: a review and classification of study designs. *Br J Radial* 1987; 60: 1071-1081.
22. Weedon D, Bastian H, Blair R, et al. Report of the Commonwealth review of positron emission tomography. http://www.health.gov.au/msac/pdfs/msacref02.pdf (accessed Dec 19, 2001).
23. Vansteenkiste JF, Stroobants SG, Dupont PJ, et al. FDG-PET scan in potentially operable non-small cell lung cancer: do anatometabolic PET-CT fusion images improve the localisation of regional lymph node metastases?. *Eur J Nucl Med* 1998; 25: 1495-1501.
24. Naruke T, Suemasu K, Ishikawa S. Lymph node mapping and curability at various levels of metastasis in resected lung cancer. *J Thorac Cardiovasc Surg* 1978; 76: 832-839.

25. International Union Against Cancer. In: TNM classification of malignant tumours. Berlin: Springer-Verlag, 1987: 69-73.
26. In: DeVita VT, Hellman S, Rosenberg SA, eds. Cancer principles and practice of oncology. Philadelphia: Lippincott-Raven Publishers, 1997.
27. Fleiss JL, Tytun A, Ury SHK. A simple approximation for calculating sample sizes for comparing independent proportions. *Biometrics* 1980; 36: 343-346.
28. In: Visser O, Coebergh JWW, Schouten LJ, van Dijck JAAM, eds. Incidence of cancer in the Netherlands, 1996. Utrecht: VIKC, 2000:.
29. Hoffman PC, Mauer AM, Vokes EE. Lung cancer. *Lancet* 2000; 355: 479-485.
30. Weder W, Schmid RA, Bruchhaus H, Hillinger S, von Schulthess GK, Steinert HC. Detection of extrathoracic metastases by positron emission tomography in lung cancer. *Ann Thorac Surg* 1998; 66: 886-892.
31. Valk PE, Pounds TR, Hopkins DM, et al. Staging non-small cell lung cancer by whole body positron emission tomographic imaging. *Ann Thorac Surg* 1995; 60: 1573-1581.
32. Van Tinteren, Hoekstra OS, Smit EF, Verboom P, Boers M. Toward less futile surgery in non-small cell lung cancer: a randomized clinical trial to evaluate the cost-effectiveness of positron emission tomography. *Control Clin Trials* 2001; 22: 89-98.
33. Vansteenkiste JF, Stroobants SG, De Leyn PR, et al. Lymph node staging in non-small-cell lung cancer with FDG-PET scan: a prospective study on 690 lymph node stations from 68 patients. *J Clin Oncol* 1998; 16: 2142-2149.
34. Kosuda S, Ichihara K, Watanabe M, Kobayashi H, Kusano S. Decision-tree sensitivity analysis for cost-effectiveness of chest 2-fluoro-2-D-[18F]fluorodeoxyglucose positron emission tomography in patients with pulmonary nodules (non-small cell lung carcinoma) in Japan. *Chest* 2000; 117: 346-353.

CHAPTER 6

Cost-effectiveness of FDG-PET in staging non-small cell lung cancer: the PLUS study

Paul Verboom[1], Harm van Tinteren[2], Otto S. Hoekstra[7, 8], Egbert F. Smit[3], Jan H. A. M. van den Bergh[4], Ad J. M. Schreurs[5], Roland A. L. M. Stallaert[6], Piet C. M. van Velthoven[4], Emile F. I. Comans[7], Fred W. Diepenhorst[2], Johan C. van Mourik[9], Pieter E. Postmus[3], Maarten Boers[8], W. M. Grijseels[1], Gerrit J. J. Teule[7], Carin A. Uyl-de Groot[1] and the PLUS study group

Institute for Medical Technology Assessment[1], Erasmus Medical Centre/Erasmus University Rotterdam, The Netherlands; Comprehensive Cancer Centre Amsterdam[2], The Netherlands; Departments of Pulmonology[3], Nuclear Medicine[7], Clinical Epidemiology and Biostatistics[8], Vrije Universiteit Medical Centre, Amsterdam, The Netherlands
Department of Pulmonology[4], Medical Centre Alkmaar, The Netherlands; Department of Pulmonology[5], Onze Lieve Vrouwe Gasthuis, Amsterdam, The Netherlands; Department of Pulmonology[6], Westfries Gasthuis, The Netherlands
A list of other participants from the PLUS Study group is shown in the Appendix.

Eur J Nucl Med Mol Imaging (2003) 30:1444-1449

Abstract

Currently, up to 50% of the operations in early-stage non-small cell lung cancer (NSCLC) are futile owing to the presence of locally advanced tumour or distant metastases. More accurate pre-operative staging is required in order to reduce the number of futile operations. The cost-effectiveness of fluorine-18 fluorodeoxyglucose positron emission tomography (^{18}FDG-PET) added to the conventional diagnostic work-up was studied in the PLUS study. Prior to invasive staging and/or thoracotomy, 188 patients with (suspected) NSCLC were randomly assigned to conventional work-up (CWU) and whole-body PET or to CWU alone. CWU was based on prevailing guidelines. Pre-operative staging was followed by 1 year of follow-up. Outcomes are expressed in the percentage of correctly staged patients and the associated costs. The cost price of PET varied between € 736 and € 1,588 depending on the (hospital) setting and the procurement of ^{18}FDG commercially or from on-site production. In the CWU group, 41% of the patients underwent a futile thoracotomy, whereas in the PET group 21% of the thoracotomies were considered futile (P = 0.003). The average costs per patient in the CWU group were € 9,573 and in the PET group, € 8,284. The major cost driver was the number of hospital days related to recovery from surgery. Sensitivity analysis on the cost and accuracy of PET showed that the results were robust, i.e. in favor of the PET group. The addition of PET to CWU prevented futile surgery in one out of five patients with suspected NSCLC. Despite the additional PET costs, the total costs were lower in the PET group, mainly due to a reduction in the number of futile operations. The additional use of PET in the staging of patients with NSCLC is feasible, safe and cost saving from a clinical and from an economic perspective.

Introduction

Lung cancer is the leading cause of cancer-related death in Western society. In 1997 in the Netherlands, approximately 8,800 people were diagnosed with lung cancer and 8,650 patients died from the disease.[1] The 5-year survival rate was 13%.[1] The poor prognosis of lung cancer can be ascribed to the high rate of unresectable disease at diagnosis and the failure of other treatment modalities to cure metastatic disease. Non-small cell lung cancer (NSCLC), for which surgery is the only curative treatment option, represents approximately 85% of all primary lung tumours.

Diagnostic work-up is essential to confirm the diagnosis of NSCLC pre-operatively and to detect (mediastinal) lymph node involvement and distant metastasis. The aim is to restrict surgery to patients who will potentially benefit.

International guidelines have been formulated to optimize these efforts, but daily clinical practice remains variable.[2] The current diagnostic work-up cannot prevent futile surgery in up to 50% of patients with NSCLC.[3] Recently, the situation has improved significantly owing to the introduction of positron emission tomography with fluorine-18 fluorodeoxyglucose (^{18}FDG-PET).[4] Results of a randomized clinical trial showed that PET, employed in addition to conventional diagnostic work-up, could reduce futile operations by 50%.[5]

The clinical relevance of PET in NSCLC is beyond doubt; however, health policy makers are concerned about the extra costs accruing from the introduction of this new technology. PET is a relatively expensive technique due to the high costs of acquisition, maintenance and the radioactive tracer.[6,7] A number of economic evaluation studies have examined PET in various diagnostic pathways through modeling. Most of these variants showed that PET could be cost-effective in NSCLC when added to conventional work-up (CWU). However, all of these studies had to rely on many assumptions and were unable to consider how cumulative diagnostic patient information impacts on patient management and final outcomes.

In the PET in Lung Cancer Staging (PLUS) study, the effectiveness of PET added to the CWU compared with CWU alone was studied in a randomized fashion consistent with routine clinical practice.[5] A cost-effectiveness analysis was anticipated in this trial. The results of this analysis are presented here. The perspective of the study was the hospital's point of view, i.e. all hospital costs associated with PET and CWU were considered.

Materials and methods

Clinical study

Between January 1998 and January 1999, patients with suspected or proven NSCLC, considered to be medically operable and to have potentially resectable disease by the referring physician on the basis of clinical staging, but not surgical staging, were invited to participate. Patients were randomly assigned either to PET followed by CWU (CWU+PET) or to CWU alone. CWU was based on existing guidelines and involved further invasive diagnostic and therapeutic procedures. Eight general hospitals and one university hospital recruited patients for the study.

PET scans were performed using a Siemens ECAT EXACT HR+ scanner (Siemens/CTI, Knoxville, Tenn., USA) at the Vrije Universiteit, Amsterdam. Referring clinicians were free to use the information and to act accordingly. However, potential distant or nodal metastases, which might have a major impact on patient management, had to be confirmed by other techniques. Unconfirmed PET findings were to be ignored. All procedures other than PET, including therapy and follow-up, were performed in the referring hospitals according to prevailing local standards. Follow-up consisted of regular visits (every 2-3 months) to the outpatient clinic for 12 months. The clinical paper contains a detailed description of the procedure.[5]

Procedure	Cost (range)
X-ray	35 (33-35)
Abdominal ultrasound	57 (24-74)
Bone scan	199 (170-260)
CT	123 (92-175)
MRI	229 (169-290)
Bronchoscopy[a]	394 (349-440)
VATS[b]	244 (202-333)
Mediastinoscopy[b]	400 (383-433)
Surgery	1,408 (1,275-1,797)
Hospital day (general ward)	220 (187-216)
ICU hospital day	1,080 (898-1,238)

VATS, Video-assisted thoracic surgery
[a]Bronchoscopy was performed with an inflexible bronchoscope, and was therefore performed in an operation room
[b]Cost due to inpatient stay was not included in the cost of the procedure, but was part of the cost of hospital days

Table 1. Cost of the various procedures (in 1999 euros)

Cost study

We focused on the costs of diagnostic and therapeutic strategies. The cost items considered were hospital days, thoracotomies, and invasive and non-invasive diagnostic tests (including PET). Therapeutic interventions such as chemotherapy and radiotherapy were not taken into account.

The full costs of the various diagnostic procedures, hospital days and intensive care days and operation costs were calculated by using the bottom-up method.[8,9] These costs included costs for personnel, materials, depreciation and overheads based on 1999 prices. The cost price study was performed in all hospitals. Where individual hospitals were not able to deliver details on specific procedures, average costs of similar procedures at the other hospitals were used. Table 1 contains the average cost prices for relevant procedures with ranges.

The costs of PET encompass the personnel costs, the depreciation and maintenance costs, the material (tracer) costs and overheads. A PET study requires a minimum level of radioactive activity, and because ^{18}FDG has a half-life of 110 min, the quantity supplied depends on the period between production and usage in the patient. The PET centre was provided twice a day with enough ^{18}FDG for the scheduled patients. The production and supply of ^{18}FDG is therefore constrained by two factors, a maximum that can be injected in the patient and a minimum level of radioactivity necessary for PET to work.

The costs of PET also depend on the daily number of scans, the average time per scan, and the amount and expertise (e.g. salary) of the personnel involved.

Several scenarios were considered by varying the number of PET scans performed per day and varying the hospital setting. The "*expensive*" variant involved a PET study performed in a university hospital with on-site tracer production and with clinical (diagnostic) as well as research functions. In this calculation, it was assumed that eight PET scans were performed daily, requiring the presence of a nuclear medicine technologist and a nuclear medicine physician. This variant was based on real costs in one centre.

The next two variants were based on expert opinion, because in the Netherlands there were only two university hospitals with a PET centre at the time of our study. In the "*cheap*" variant, 12 PET scans per day were performed in a community hospital with a PET scanner which was incorporated in a nuclear medicine department, and merely used for clinical diagnostic work. Production and transport of ^{18}FDG were done according to the satellite concept.[7] The "*in-between*" option considered a large community hospital in which eight PET scans were performed per day. In this setting, limited research was done and ^{18}FDG was produced on site.

Table 2 shows a detailed specification of the cost of a PET scan. In this study we used the "in-between" variant. In this variant the costs of a whole-body FDG-PET scan amounted to € 1,020. The upper and lower values were used in a sensitivity analysis.

	Cheap variant	"In between" variant	Expensive variant[a]
Personnel	160	208	570
FDG/material	227	245	259
Depreciation/maintenance	227	302	481
Overhead	123	264	277
Total	736	1,020	1,588

[a]Including research

Table 2. Cost of a PET scan in different scenarios

Outcomes

Primary study outcome was the number of futile thoracotomies. A thoracotomy was classified as futile in cases of: (1) a benign lung lesion, (2) pathologically proven mediastinal lymph node (MLN) involvement (stage IIIA-N2) other than minimal N2 disease (i.e. intranodal involvement in a single lymph node established at mediastinal dissection), (3) stage IIIB disease, (4) explorative thoracotomy for any other reason, or (5) recurrent disease or death from any cause within 1 year after randomization.

Patients were analyzed according to the intention-to-treat principle. Secondary outcome was the cost associated with both strategies calculated as the average cost per (suspected) NSCLC patient.

Sensitivity analysis

Sensitivity analyses were performed on the efficacy of PET and the PET setting. For the efficacy the 95% confidence interval was used and for the setting the "expensive" variant and the "cheap" variant were taken.

Statistical analyses

The number of futile thoracotomies was tested by the chi-square test based on intention to treat. The costs of the various cost items were tested by means of the two-tailed Wilcoxon-Mann-Whitney test.

Results

Clinical results

Table 3 presents the baseline characteristics. Seventy-one percent of the CWU patients and 70% of the CWU+PET patients had clinical stage I or II. Pre-randomization work-up was similar for both groups and included at least a chest computed tomography (CT) scan, which included the liver and adrenals (89%). In both groups, 58% of patients underwent at least one additional test to identify metastatic disease before randomization.

	CWU (*n*=96)	CWU+PET (*n*=92)
Mean age (years, SD)	65 (10)	66 (10)
Male sex, *n*	75 (78%)	69 (75%)
Karnofsky index		
70-80	6 (6%)	10 (11%)
90-100	90 (94%)	82 (89%)
Weight loss >5%	15 (16%)	14 (15%)
Clinical stage		
I	63 (66%)	58 (63%)
II	5 (5%)	6 (7%)
IIIA	22 (23%)	23 (25%)
IIIB	6 (6%)	4 (4%)
IV	0	1 (1%)

Table 3. Baseline characteristics, clinical study

The primary outcome is the proportion of patients who underwent futile thoracotomy: in the CWU arm, 41% (39/96) underwent futile surgery compared with only 21% (19/92) in the CWU+PET arm (P=0.003) (Table 4). This implies a relative reduction in the number of futile surgical procedures by 51%. The absolute difference of 20% corresponds to five patients (95% confidence interval 3–14) needing CWU+PET to avoid one futile thoracotomy. Eighteen patients in the CWU arm and 32 in the CWU+PET arm did not proceed to thoracotomy, mainly due to the presence of benign lesions, upstaging due to detection by PET or intercurrent morbidity and recurrence of metastases. Eighteen percent of the CWU patients and 32% of the CWU+PET patients were not operable owing to extensive co-morbidity or refusal of the patient. Unfortunately, in both arms several patients died due to surgery-related complications. Apparently justified surgery was done in a closely similar number of patients in each group, namely 39/96 in the CWU group and 41/92 in the CWU+PET group.

	CWU (*n*=96)	CWU+PET (*n*=92)
Thoracotomy	78 (81%)	60 (65%)
Non-futile surgery	39 (41%)	41 (44%)
Futile surgery	39 (41%)	19 (21%)
No thoracotomy	18 (19%)	32 (35%)

Table 4. Primary outcome of the clinical study

	CWU	CWU+PET
Operated patients		
No. (%)	78 (81%)	60 (65%)
Mean no. of (non-)invasive tests	1.6	1.8
Mean no. of general hospital days (SD)	19.8 (18.6)	14.9 (10.1)
Mean no. of IC days (SD)	4.7 (9.3)	4.0 (7.1)
Non-operated patients		
No. (%)	18 (19%)	32 (35%)
Mean no. of (non-)invasive tests	1.8	1.9
Mean no. of general hospital days (SD)	4.5 (6.7)	8.6 (13.6)
Mean no. of IC days (SD)	0.1 (0.24)	0.3 (1.4)
All patients		
No.	96	92
Mean no. of (non-)invasive tests	1.6	1.9
Mean no. of general hospital days (SD)	16.9 (18.0)	12.7 (11.7)
Mean no. of IC days (SD)	3.8 (8.5)	2.7 (6.1)

Table 5. Most important cost items

Cost analysis

Considering the groups of operative patients and all patients, the mean number of regular hospital days and number of days on the intensive care ward were in favor of the CWU+PET group. The number of imaging tests and (non-)invasive tests did not differ between the groups (except for the PET scan). Also the number of mediastinoscopies was similar in the two arms (63 patients in the CWU group and 67 patients in the CWU+PET group). The most important cost items are shown in Table 5.

The group of patients who were not operated on consisted of patients who refused an operation or patients who had severe co-morbidity (six patients), were down- or upstaged (41 patients) or were discovered to have a tumor type for which surgery was not the optimal treatment (three patients). Overall, the average total costs were lower in the PET group owing to a reduction in (futile) operations and subsequent hospital days (Table 6). Patients who underwent surgery in the CWU arm cost € 11,486 on average, versus € 10,709 in the CWU+PET arm. Patients who underwent futile operations in the CWU group cost on average € 12,473, compared with € 13,689 in the CWU+PET group. This difference is mainly related to the cost of PET (€1,021) in the CWU+PET arm. Remarkably, non-futilely operated patients cost on average € 10,489 in the CWU group and € 9,487 in the CWU+PET group. This difference was mainly caused by a higher number of intensive care days due to one patient in the CWU group, who stayed 61 days on the intensive care ward.

	CWU	CWU+PET
Operated patients		
Imaging tests	–	13 (51)
(Non-)invasive tests	279 (193)	288 (191)
PET	–	970[a] (224)
Surgery[b]	1,576 (359)	1,637 (425)
General hospital ward	4,543 (4,177)	3,439 (2,260)
IC ward	5,088 (10,099)	4,362 (7,744)
Total costs (median; SD)	11,486 (8,020; 12,628)	10,709 (8,621; 7,727)
Non-operated patients		
Imaging tests	2 (8)	38 (69)
(Non-)invasive tests	226 (208)	318 (171)
PET	–	1,021 (0)
Surgery[b]	–	–
General hospital ward	999 (1,466)	1,987 (3,222)
IC ward	60 (255)	373 (1,524)
Total costs (median; SD)	1,287 (416; 1,609)	3,736 (2,521; 4,137)
All patients		
Imaging tests	0	22 (59)
(Non-)invasive tests	269 (196)	298 (184)
PET	0	988 (182)
Surgery[b]	1,281 (698)	1,068 (855)
General hospital ward	3,879 (4,057)	2,934 (2,707)
IC ward	4,145 (9,304)	2,975 (6,582)
Total costs (median; SD)	9,573 (7,480; 12,072)	8,284 (7,592; 7,462)

SD, Standard deviation
[a] Two patients refused PET imaging
[b] Including re-intervention surgery due to complications

TABLE 6. Average costs (SD) of CWU and CWU+PET (costs in Euros)

The average costs of patients who were not operated on amounted to € 1,287 in the CWU group and to € 3,851 in the CWU+PET group. The costs in the CWU group were lower because the patients spent fewer days in the hospital and because of the cost of PET itself.

Sensitivity analysis

The results of the sensitivity analysis are presented in Table 7.

Setting/cost of PET	No. of futile operations remaining after PET			
	8	19	26	36
Cheap variant: Euro 736	−2,401	−1,565	−1,033	−274
"In between" variant: Euro 1,020	−2,123	−1,289	−759	0[a]
Expensive variant: Euro 1,588	−1,571	−741	−212	+542

[a] Break-even point

Table 7. Sensitivity analysis on the efficacy and the setting of PET: difference compared with cost of CWU (cost of CWU = Euro 9,573)

Varying the efficacy of PET

In this study adding PET to CWU resulted in a 20% absolute reduction (from 41% to 21%) in futile operations or, put another way, one prevented unnecessary surgical procedure for every five PET scans (95% CI: 3–14). The 95% confidence interval corresponded to a range of one prevented unnecessary surgical procedure for every three PET scans (9% futile operations remaining in the PET group) to one prevented procedure for every 14 PET scans (28% futile operations remaining). Taking into account the cost of €1,021 per PET scan, the more efficient PET result would result in higher savings compared with CWU. Preventing 28% of the futile operations would result in a difference of €2,123 in favor of CWU+PET. Even the upper boundary of the 95% CI would result in a difference of €759 in favor of CWU+PET.

Varying the setting of PET

The results of the cost analysis were rather sensitive to the price of PET. If the PET costs were lower than €1,021, the savings in favor of the PET+CWU arm would increase. Furthermore, the PET costs in the "expensive" setting were still in favor of the PET+CWU group. The break-even point, i.e. with no difference in costs between the groups, was at a PET cost of €2,350. If the costs of PET were higher than €2,350, then the CWU group would be cheaper.

Varying the setting and the efficacy of PET

Varying both the setting and the efficacy of PET, most results were robust and in favor of the PET+CWU group. When taking into account the highest price of PET and the worst efficacy outcome, the results were in favor of the CWU arm (namely €542).

Discussion

The introduction of PET as a new diagnostic technology for various indications is clinically an important step, but the costs may be relatively high. In this randomized controlled trial we directly measured the efficacy of PET as well as the associated costs. Our study shows that PET added to CWU in patients with suspected NSCLC is effective from either perspective. Since hospital days, the operation itself and postoperative intensive care were the major cost drivers, a decrease in futile operations would have a profound effect on the costs.

Until now, cost-effectiveness data on PET could only be derived from studies that applied modeling techniques to estimate the costs and benefits of PET in various diseases,[10-17] and mainly in (suspected) lung cancer. The results of these studies have been based on reviews of the sensitivity and specificity of CT and PET. Morbidity and mortality effects have scarcely been reported, and have been reliant upon additional assumptions made by the investigators. Diagnostic accuracy measures do not easily translate to the complete clinical decision-making and management of patients and subsequent clinical outcome. Further, one of the main deficiencies of these analyses is a failure to consider how cumulative diagnostic information impacts on patient management and final outcomes.[15] However, improved accuracy in pre-operative staging should provide an improvement in survival.[18]

A limitation of our study is the neglect of pre-operative or neoadjuvant treatments for patients with locally advanced NSCLC or other treatment options for patients who were unsuitable for thoracotomy. However, it has been claimed that multimodality interventions are cost-effective.[6]

The real costs of a PET scan are complicated to assess. In the Netherlands, the application is relatively new and no guidelines regarding indications for PET have been established at the national level. Furthermore, the acquisition cost of a PET scanner varies between €1,000,000 and €2,800,000, so that the depreciation per scan can vary considerably. The daily production of PET scans is an important determinant of the calculated price of a PET scan. Since clinical PET may comprise research as well as diagnostic activities, the daily production of diagnostic scans may range from 6 to 12. Therefore the cost of PET shows considerable variation within the hospital setting.

The possible configurations for PET and cyclotron have also been described elsewhere. In the United States, Keppler et al. found a range between $2,986 and $1,557 (i.e. between €2,706 and €1,411) (including radioactive FDG) for a whole-body PET scan.[13,14] In our study, the break-even cost of PET (in which the average costs per NSCLC patient would have been equal in both arms) was €2,234, . This cost price of PET was much higher than the cost in the most expensive scenario (i.e. a university hospital with a research function). In this scenario the cost price amounted to €1,588.

In some countries, PET scans are reimbursed by insurance companies. In the USA, reimbursement amounts to approximately $2,000 (^{18}FDG included) for PET staging in lung cancer. In Germany the reimbursement amounts to €1,227.[11] In the Netherlands, the results of this study could be used for the reimbursement policy. Here, about 9,000 new lung cancers are detected annually. About 85% of these are NSCLC. After exclusion of those patients who prove to have irresectable disease after standard dissemination tests and those who are inoperable owing to co-morbidity precluding sur-

gery, approximately 5,000 patients would be eligible for PET. Applying the results of this trial, the routine use of PET for NSCLC could save a great number of thoracotomies, resulting in a decrease in morbidity and resource use (i.e. intensive care), with a consequent saving of around 6.4 million euros (5,000 × €1,289).

Recently a second randomized controlled trial of PET in NSCLC was closed after including a total of 470 patients. The primary research question of this study is whether PET may substitute for other techniques in the conventional work-up. Again, data on costs will be collected in detail and analyzed as a secondary endpoint.

On the basis of this study we conclude that the additional use of PET in the staging of patients with NSCLC is feasible and safe and saves costs from a clinical and an economic perspective.

Appendix

In addition to the authors, the PLUS Study Group participants were as follows: A. Boonstra, B. Venmans, A.J.M. van Boxem, G. Sutedja, R.A. Manoliu and R. Golding at the Academic Hospital Vrije Universiteit, Amsterdam; W.F.M. Strankinga, P.M. Hooghiemstra, F.C. Crezée and P.L. Tolenaar at the BovenIJ Ziekenhuis, Amsterdam; H.B. Kwa, J.G. van Unnik, R. Hardjowijono and S.S. Wagenaar at the Onze Lieve Vrouwe Gasthuis, Amsterdam; C. Jie, M.C.T.B. Sie, R.M.J.M. Butzelaar and M.N. Weimann at the Lucas Ziekenhuis, Amsterdam; G. Visschers, P.I. van Spiegel, G.C. Collet and R.P. Rademakers at the Slotervaart Ziekenhuis, Amsterdam; W.G. Boersma, M. Deenstra, C.S. de Graaff, T. Haitjema, G.H. Ooms and J.H. Pot at the Medisch Centrum Alkmaar; J. Prins, P.M.J.M. de Vries, E. Scheijde and J. Dijkstra at the Westfries Gasthuis, Hoorn; J.P. Teengs and E. Boerma at the Kennemer Gasthuis, Haarlem; and J. Berkovits, R.P.A. Boom and J Krekt at the Ziekenhuis Amstelveen.

References

1. Jansen-Heynen MLG, van Dijck JAAM, Schipper RM, Damhuis RAM. Longkanker in Nederland in de periode 1989–1997: de epidemie is nog niet voorbij. *Ned Tijdschr Geneeskd* 2001; 145:419–423.
2. Herder GJM, et al. Staging of non-small cell lung cancer in two large Dutch hospitals. *Eur J Respir Dis* 1999; 14:446.
3. van Tinteren H, Hoekstra OS, Smit EF, Verboom P, Boers M. Towards less futile surgery in non-small cell lung cancer? A randomised clinical trial to evaluate cost-effectiveness of positron emission tomography. *Control Clin Trials* 2001; 22:89–98.
4. Pieterman RM, van Putten JWG, Meuzelaar JJ, et al. Preoperative staging of non-small cell lung cancer with positron-emission tomography. *N Engl J Med* 2000; 343:254–261.
5. van Tinteren H, Hoekstra OS, Smit EF, et al. Effectiveness of positron emission tomography in the preoperative assessment of patients with suspected non-small-cell lung cancer: the PLUS multicentre randomised trial. *Lancet* 2002; 359:1388–1393.
6. Evans WK, Will BP, Berthelot JM, Earle CC. Cost of combined modality interventions for stage III non-small cell lung cancer. *J Clin Oncol* 1997; 15:3038–3048).
7. Lottes G, Gorschlutter P, Kuwert T, Adam D, Schober O. Costs of F-18-FDG-PET. Experience with a satellite concept. *Nuklearmedizin* 1998; 37:159–164.
8. Verboom P, Herder GJM, Hoekstra OS, et al. Staging of non-small cell lung cancer and application of FDG-PET: a cost approach. *Int J Technol Assess Health* C 2002; 18:576–585.
9. Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-effectiveness in health and medicine. New York: Oxford University Press, 1996.
10. Gambhir SS, Hoh CK, Phelps ME, Madar I, Maddahi J. Decision tree sensitivity analysis for cost-effectiveness of FDG-PET in the staging and management of non-small cell lung carcinoma. *J Nucl Med* 1996; 37:1428–1436.
11. Dietlein M, Weber K, Gandjour A, et al. Cost-effectiveness of FDG-PET for the management of solitary pulmonary nodules: a decision analysis based on cost reimbursement in Germany. *Eur J Nucl Med* 2000; 27:1441–1456.
12. Keith CJ, Miles KA, Griffiths MR, et al. Solitary pulmonary nodules: accuracy and cost-effectiveness of sodium iodide FDG-PET using Australian data. *Eur J Nucl Med Mol Imaging* 2002; 29:1016–1023.
13. Keppler JS. Federal regulations and reimbursement for PET. *J Nucl Med Technol* 2001; 29:173–179; quiz 180–182.
14. Keppler JS, Conti PS. A cost analysis of positron emission tomography. *Am J Roentgenol* 2001; 177:31–40.
15. Miles KA. An approach to demonstrating cost-effectiveness of diagnostic imaging modalities in Australia illustrated by positron emission tomography. *Australas Radiol 2001*; 45:9–18.
16. Scott WJ, Shepherd J, Gambhir SS. Cost-effectiveness of FDG-PET for staging non-small cell lung cancer: a decision analysis. *Ann Thorac Surg* 1998; 66:1876–1885.
17. Dietlein M, Weber K, Gandjour A, et al. Cost-effectiveness of FDG-PET for the management of potentially operable non-small cell lung cancer: priority for a PET-based strategy after nodal-negative CT results. *Eur J Nucl Med* 2000; 27:1598–1609.
18. Vesselle H, Pugsley JM, Vallieres E, Wood DE. The impact of fluorodeoxyglucose F 18 positron-emission tomography on the surgical staging of non-small cell lung cancer. *J Thorax Cardiovasc Surg* 2002; 124: 511–519.

CHAPTER 7

Controversies – For

Do we need randomized trials to evaluate diagnostic procedures?

H. Van Tinteren[1], O.S. Hoekstra[2, 3] and M. Boers[3]

Comprehensive Cancer Centre Amsterdam[1], Plesmanlaan 125, 1066 CX Amsterdam, The Netherlands; Department of Nuclear Medicine/Clinical PET Centre[2], Department of Clinical Epidemiology and Biostatistics[3] VU University Medical Centre, Amsterdam, The Netherlands

Eur J Nucl Med Mol Imaging (2004) 31:129-132

The VU Medical Centre serves a population of 2.5 million people and so, in 1996, when it was decided that the Centre would acquire a dedicated positron emission tomography (PET) scanner, it was anticipated that diagnosticians and clinicians would soon be confronted with a lack of scan capacity. Decisions would need to be made on which clinical indications and, within those indications, at what point in the diagnostic work-up PET would be most beneficial. Our main concern was to try to ensure equity of access to this new technology. We insisted that routine use of PET should be restricted to those applications where it would have a clinically meaningful impact, reasoning that waiting lists are not compatible with routine patient care. In cancer care, most of the literature on the use of PET focused on non-small cell lung cancer (NSCLC), which was also the most common cancer in terms of incidence. Moreover, a retrospective study in our region identified a relevant clinical problem in lung cancer diagnosis.[1]

Although accuracy studies showed that PET improved diagnostic performance, it was unclear how and to what extent these results might affect management and patient outcome. Direct evidence did not exist on whether the addition of PET to conventional staging in daily clinical practice would lead to a clinically relevant reduction of unnecessary operations for patients with suspected resectable NSCLC. We decided that randomized trials would provide the most reliable and convincing answer to this uncertainty. Randomization provides broad comparability of groups except for the diagnostic device being assessed and thereby ensures a non-biased assessment of the value of the experimental technique with respect to clinically relevant outcomes. Although randomization in diagnostic research is slowly becoming more accepted, it appears that the value of randomized trials is still disputed.[2-5] All aspects that used to be seen as challenges to the use of randomization for assessing medical interventions in the 1970s, such as the difficulties of performing randomized trials, their adequacy and their conclusiveness, are now being raised as arguments against randomized trials of diagnostic techniques.[6]

This paper describes practical situations in which randomization would be feasible and efficient. These arguments will be followed by a discussion on methodological issues that need special consideration.

Accuracy and beyond

In NSCLC, evaluation of the extent of spread in the mediastinum is an important step in deciding about further therapy. In general, patients with pre-operative identifiable N2 disease are excluded from thoracotomy. Conventional work-up of the mediastinum often includes non-invasive imaging with computed tomography (CT) followed by invasive exploration through mediastinoscopy. In a recently published meta-analysis of 18 studies, including more than a thousand patients, the negative predictive value of PET was found to be 93%, which is at least as good as that of mediastinoscopy in daily clinical practice.[7] However, improved diagnostic performance does not necessarily translate into meaningful changes in clinical decisions, even at the level of this relatively simple diagnostic problem. In fact, in spite of the burgeoning number of PET studies, the 2003 guidelines on NSCLC staging[8] do not clearly recommend that mediastinoscopy can be omitted if a PET scan is negative. Moreover, in general it will be impossible to measure impact on clinical outcomes in accuracy

studies, because the design of such studies requires that the new imaging technique is assessed out of the clinical context. For example, clinically relevant PET findings (such as the suspicion of a distant metastasis in an otherwise resectable cancer) need to be confirmed because of the risk that this might be a false positive finding.[9] Since this is not allowed in the context of an accuracy study on mediastinal staging it needs to be proven that such confirmation is feasible in daily clinical practice.[10] Further, in clinical practice compared with accuracy studies, it is less common to have dichotomous test results,[11,12] and grey areas prevail with grading of diagnostic suspicion arising from a variety of test results. This contributes to the difficulty of assessing how the technique will perform in clinical practice when it is used in combination with other diagnostic interventions. How will the clinician use the newly provided diagnostic information and how will this affect the performance of other investigations and treatment?

In many ways, the level of information provided by accuracy studies can be compared to that from phase II studies in the evaluation of treatments. These studies are a prerequisite in the process of building evidence of activity, but not suited to producing information on the difference between the effects of a new drug and those of other competing drugs or devices in routine practice.[13]

Alternatively, decision analyses can be used to estimate the effect of various strategies on cost and effectiveness outcomes. Because the input measures are usually obtained from accuracy studies that often do not comply with quality standards[14-16] and assumptions are required to make the problem tractable,[5] this technique is also unsuitable when faced with the complexity of daily clinical practice and the need to make decisions with respect to clinically meaningful outcomes. Nevertheless, in the presence of many alternative diagnostic strategies, such an approach may be helpful in identifying the most promising diagnostic tests or algorithms for further research.[17]

The extent to which a patient may ultimately benefit from the addition of a new imaging technique (in terms of reduction in iatrogenic toxicity or improvement in survival) can only be investigated through a comparison of the full implementation of PET added to conventional staging versus the conventional process alone. The benefit of the diagnostic interventions will often be no more than moderate, just like the experiences with therapeutic interventions. In such circumstances, then, just as with treatments, it is important that the comparison is done in such a way that both systematic and chance effects are minimized as much as possible. Balancing both known and unknown prognostic variables by randomly assigning some patients to the new test and others to the control group is the most efficient way to do this, along with making the study as large as possible. The process of random assignment will mean that the diagnostic test should be the only difference between the two groups. As well as minimizing imbalances in prognostic factors, randomized studies also have several qualities and benefits that arise not from the act of random allocation itself, but from the fact that they have many features of high-quality research.[18] A written protocol provides transparency, the pre-calculation of a sample size often allows some exploration of patient and tumor characteristics that might be outcome related, and direct cost comparisons are also possible.

We conclude that randomized studies are important in advancing diagnostic technology beyond the level of accuracy. However, several issues bear consideration.

Outcome measures

Conventionally, the outcome of most randomized studies is measured in terms of patient mortality and morbidity. However, diagnostic tests serve to allocate appropriate management to patients. Thus, as long as the new test does not alter the definitions of staging and management for each stage is already established, a reasonable outcome measure for such studies will be the extent to which appropriate therapy is applied. With NSCLC, there was a broad agreement among clinicians that treatment options would be clear when the diagnostic process had been completed. In addition, because PET did not produce a new stage classification, a suitable outcome measure could be the reduction of iatrogenic morbidity (translated into unnecessary operations) rather than survival.[19]

However, in some situations, new tests may identify new prognostic subsets. Then, the result of the test might affect both the diagnosis and the intervention. A randomized trial with mortality-related outcomes would then be needed to assess the effect of the test. We will illustrate this with two examples. The sentinel node biopsy (SN) in breast cancer was developed merely to reduce the number of unnecessary axillary dissections. Later, it became clear that the procedure also induced stage migration: the single biopsied (sentinel) node allowed a much more thorough histopathological evaluation than had been feasible with the 15-20 nodes typically harvested from an axillary specimen of a woman with breast cancer. This meant that prognostic and, therefore, treatment issues might be affected: implementation of the SN biopsy might unintentionally lead to overtreatment of women if they received adjuvant therapy from which they could derive no benefit. This is close to the notion of a "false positive sentinel node".

In the evaluation of non-Hodgkin's lymphoma, fluorine-18 fluorodeoxyglucose (FDG) PET discriminates responders and non-responders much better than the current CT strategy.[20] It might be expected, therefore, that this would lead to improved survival because of the earlier recognition of therapy failure (by PET compared with clinical follow-up) and the initiation of second-line therapy. However, to assess this reliably requires a direct, randomized comparison of FDG-PET versus routine management.

In summary, the choice of the outcome should be considered carefully with respect to the potential effect of the diagnostic process on stage and management, but it is not necessarily restricted to survival.

Applicability

As with randomized trials of treatments, the most valid estimate of the (cost) effectiveness of a diagnostic test is likely to come from multiple multicenter trials. However, especially in large pragmatic trials, the variability in the control groups may be larger than that found in trials of therapy, and the added value of the new technique relates to the yield from the standard procedures, without the new test. In a setting in which the outcomes for patients undergoing diagnostic procedures without the new test are very different from those for the control patients in the trial, the effect found in the trial might not be applicable. Therefore, the first step in trying to apply the trial's results to a local situation is to analyze the extent and yield of current diagnostic practice in that setting. This is best done prior to introducing the new diagnostic procedure.[1]

"The times, they are a-changin'"

There is concern that technology proceeds so rapidly that a device tested in complex randomized trials may be outdated by the time that these are analysed and reported.[2] We believe that this argument is often over-stated. Since the introduction of whole-body FDG-PET in oncology, the technique has not essentially changed for at least a decade. If there had been a switch to outcome-oriented studies as soon as accuracy studies suggested clinically relevant benefit, time could have been saved.

It is only if a new technology is so obviously better and costs less, that randomized assessments of it will not be necessary. Examples of such technologies are extremely rare.

Recently, integrated PET-CT scanners have been introduced. There is no doubt that image quality and ease of interpretation will strongly appeal to clinicians. An important improvement in diagnostic accuracy in NSCLC staging has already been reported.[11] If this development leads to better staging and a corresponding impact on patient management, a single-step diagnostic work-up may be within reach. The downside is that shortages of scan capacity will be even greater than we faced in 1997 with the introduction of the old PET scanner. However, whether the improved accuracy claimed for the integrated device will solve the residual clinical problems of PET visually correlated with CT scans and indeed lead to better patient outcomes remains to be demonstrated. We doubt whether improved T-staging by PET-CT fusion will make surgeons refrain from potentially curative resection. Further, to what extent will the proportion of equivocal results encountered in daily practice be reduced and will this indeed lead to improved management? Currently, with limited availability of the new technology, an ideal window of opportunity exists for randomized trials and international collaborative action should be sought to perform such studies.

In conclusion, there are good arguments in favour of randomization in diagnostic research. The results of randomized trials provide the reliable evidence base needed for clinical decision-making. The PLUS study,[19] investigating FDG-PET in NSCLC staging, shows that pragmatic randomized trials are feasible. Unless there are compelling reasons to the contrary, randomized studies should be required for new diagnostic interventions. These studies should use clinically relevant outcome measures and be conducted as early as possible in the evaluation of the new technology.

References

1. Herder GJ, Verboom P, Smit EF, van Velthoven PC, van den Bergh JH, Colder CD, van Mansom I, van Mourik JC, Postmus PE, Teule GJ, Hoekstra OS. Practice, efficacy and cost of staging suspected non-small cell lung cancer: a retrospective study in two Dutch hospitals. *Thorax* 2002; 57:11-14.
2. Hojgaard L. Are health technology assessments a reliable tool in the analysis of the clinical value of PET in oncology? Who audits the auditors? *Eur J Nucl Med Mol* Imaging 2003; 30:637-641.
3. Valk PE. Randomized controlled trials are not appropriate for imaging technology evaluation. *J Nucl Med* 2000; 41:1125-1126.
4. Dixon AK. Evidence-based diagnostic radiology. *Lancet* 1997; 350:509-512.
5. Hunink MG, Krestin GP. Study design for concurrent development, assessment, and implementation of new diagnostic imaging technology. *Radiology* 2002; 222:604-614.
6. Byar DP, Simon RM, Friedewald WT, Schlesselman JJ, DeMets DL, Ellenberg JH, Gail MH, Ware JH. Randomized clinical trials. Perspectives on some recent ideas. *N Engl J Med* 1976; 295:74-80.
7. Canadian Lung Oncology Group. Investigation for mediastinal disease in patients with apparently operable lung cancer. *Ann Thorac Surg* 1995; 60:1382-1389.
8. Silvestri GA, Tanoue LT, Margolis ML, Barker J, Detterbeck F. The noninvasive staging of non-small cell lung cancer: the guidelines. *Chest* 2003; 123 (1 Suppl):147S-156S.
9. Hoekstra CJ, Stroobants SG, Hoekstra OS, Vansteenkiste J, Biesma B, Schramel FJ, van Zandwijk N, van Tinteren H, Smit EF. The value of [^{18}F]fluoro-2-deoxy-D-glucose positron emission tomography in the selection of patients with stage IIIA-N2 non-small cell lung cancer for combined modality treatment. *Lung Cancer* 2003; 39:151-157.
10. Pieterman RM, van Putten JW, Meuzelaar JJ, Mooyaart EL, Vaalburg W, Koeter GH, Fidler V, Pruim J, Groen HJ. Preoperative staging of non-small-cell lung cancer with positron-emission tomography. *N Engl J Med* 2000; 343:254-261.
11. Lardinois D, Weder W, Hany TF, Kamel EM, Korom S, Seifert B, von Schulthess GK, Steinert HC. Staging of non-small-cell lung cancer with integrated positron-emission tomography and computed tomography. *N Engl J Med* 2003; 348:2500-2507.
12. Steinert HC, Hauser M, Allemann F, Engel H, Berthold T, von Schulthess GK, Weder W. Non-small cell lung cancer: nodal staging with FDG-PET versus CT with correlative lymph node mapping and sampling. *Radiology* 1997; 202:441-446.
13. Freedman LS. Evaluating and comparing imaging techniques: a review and classification of study designs. *Br J Radiol* 1987; 60:1071-1081.
14. Toloza EM, Harpole L, McCrory DC. Noninvasive staging of non-small cell lung cancer: a review of the current evidence. *Chest* 2003; 123 (1 Suppl):137S-146S.
15. Gould MK, Maclean CC, Kuschner WG, Rydzak CE, Owens DK. Accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions: a meta-analysis. *JAMA* 2001; 285:914-924.
16. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HC. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Radiology* 2003; 226:24-28.
17. Gould MK, Sanders GD, Barnett PG, Rydzak CE, Maclean CC, McClellan MB, Owens DK. Cost-effectiveness of alternative management strategies for patients with solitary pulmonary nodules. Ann Intern Med 2003; 138:724-735.
18. Abel U, Koch A. The role of randomization in clinical studies: myths and beliefs. *J Clin Epidemiol* 1999; 52:487-497.
19. van Tinteren H, Hoekstra OS, Smit EF, van den Bergh JH, Schreurs AJ, Stallaert RA, van Velthoven PC, Comans EF, Diepenhorst FW, Verboom P, van Mourik JC, Postmus PE, Boers M, Teule GJ. Effectiveness of positron emission tomography in the preoperative assessment of patients with suspected non-small-cell lung cancer: the PLUS Multicentre Randomized Trial. *Lancet* 2002; 359:1388-1393.

20. Spaepen K, Stroobants S, Dupont P, Van Steenweghen S, Thomas J, Vandenberghe P, Vanuytsel L, Bormans G, Balzarini J, Wolf-Peeters C, Mortelmans L, Verhoef G. Prognostic value of positron emission tomography (PET) with fluorine-18 fluorodeoxyglucose ([^{18}F]FDG) after first-line chemotherapy in Non-Hodgkin's lymphoma: is [^{18}F]FDG-PET a valid alternative to conventional diagnostic methods? *J Clin Oncol* 2001; 19:414-419.

CHAPTER 8

Evaluating positron emission tomography in non-small-cell lung cancer: moving beyond accuracy to outcome

Harm van Tinteren[1], Otto S. Hoekstra[2,4], Carin A. Uyl-De Groot[3] and Maarten Boers[4]

Comprehensive Cancer Center Amsterdam[1], Amsterdam, The Netherlands; Departments of Nuclear Medicine[2], Clinical Epidemiology and Biostatistics[4], Vrije Universiteit Medical Center, Amsterdam, The Netherlands; Institute for Medical Technology Assessment[3], Erasmus Medical Center/Erasmus University, Rotterdam, The Netherlands

Adapted from *Cancer Imaging*, M.A.Hayat eds. Elsevier Academic Press, in press.
And *Clinical Oncology* 2006; 18(2):156-157

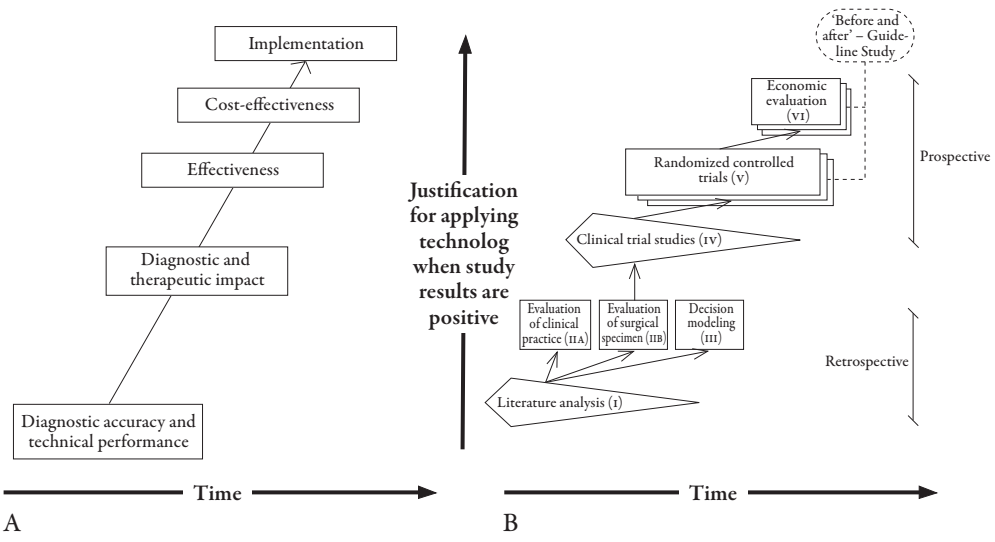
Summary

Accuracy measures of diagnostic tests usually do not allow an appropriate assessment of its (cost-)effectiveness in clinical practice. Since test results are seldom 'black or white' and are usually part of a complex work-up process, the benefit in overall patient outcome is preferably measured by comparing the diagnostic process including the new test concurrently with the prevailing diagnostic work-up. Obviously, a trial should be preceded by formal analyses of residual (in)efficiency of the prevailing diagnostic work-up and of the potential diagnostic and therapeutic impact of the new test. This hierarchical approach as described by Fryback and others is generally considered the optimal way to evaluate new and costly diagnostic imaging devices.[39] However, practical examples of this stepwise process related to a particular device or indication, especially with regard to outcome levels beyond accuracy, are scarce. This chapter presents a series of coherent studies that have been performed to evaluate the use of PET for specific disease indications in cancer with clinically relevant endpoints beyond accuracy. These studies do not require exceptional resources.

Introduction

The introduction of new technology in health care is often accompanied by a fundamental dilemma: rapid diffusion inspired by the perspective of considerable clinical benefit versus adequate evaluation followed by implementation in appropriate clinical situations. This dilemma was recognized shortly after the introduction of computed tomography (CT) in 1973 in the United States,[1] and again with the introduction of magnetic resonance imaging (MRI) a decennium later.[2] Compared to therapeutic interventions, the evaluation of a diagnostic test is particularly challenging because the results may be less directly translatable into health outcome. The value of plain radiography to diagnose bone fracture is unquestionable, and its result directly dictates patient management. However, when clinical decision-making is multifactorial and test results are imperfect, improved diagnostic accuracy provided by one component of a diagnostic test sequence may not necessarily translate into meaningful therapeutic changes. Obviously, without evidence of changes in therapeutic decisions, the implementation of new tests is unlikely to have any impact on health or costs.

A hierarchical approach to the assessment of diagnostic imaging technology has been advocated by experts in health technology assessment (Figure 1A).[3] The framework typically assumes that evaluation of technical and image quality and diagnostic performance (sensitivity and specificity)



Note: the sharp elements in figure 1B indicate that these processes may continue in time. The dotted element with rounded edges indicates a step that has not been studied in full depth.

FIGURE 1. Theoretical hierarchical approach to development, assessment and implementation of new diagnostic technology (A) and framework of study designs, beyond accuracy to outcome, applying the theoretical approach (B).

is followed by assessment of diagnostic and therapeutic impact. Then, effectiveness on patient and societal outcomes needs to be measured, and the process is completed by defining costs and benefits of the implementation. Although this approach has been cited and discussed frequently in the literature neither CT nor MRI was assessed accordingly before clinical adoption.

Typically, the first two levels were studied extensively but the implementation of the technique was independent of evidence from the levels beyond accuracy. Another 20 years later, positron emission tomography (PET) was embraced with similar excitement, but has limitations.[4] Although studies on the accuracy of PET have been improved by using methodological and reporting guidelines,[5] research to establish the role of PET in changing patient outcome and cost-effectiveness failed to appear. Current constraints on health care resources and the limited availability of the technique in our situation stimulated us to develop a program to evaluate the clinical role of ^{18}F-Fluorodeoxyglucose-PET.

The case of ^{18}F-Fluorodeoxyglucose – positron emission tomography

In 1996 the VU Medical Center (VUmc) acquired a dedicated PET scanner; at that time the second scanner in the Netherlands. The VUmc serves a population of 2.5 million people; thus a lack of scan capacity was to be anticipated. Data were urgently needed to decide for which clinical indications and, within those indications, at what point in the diagnostic work-up ^{18}FDG-PET would be most beneficial. This system is mostly suggested as an 'add-on' technique, but it might also substitute for other diagnostic procedures. Our main concern was to ensure equity of access to this new technology. To avoid waiting lists, we restricted ^{18}FDG-PET to clinical research and to those applications where sufficient evidence for a clinically meaningful impact was available.

In this chapter, first, the strengths and limitations of accuracy studies will be summarized, exemplified by the situation in non-small-cell lung cancer (NSCLC). Then, a series of coherent studies will be presented with endpoints beyond the level of diagnostic accuracy. This framework follows the hierarchical approach for the assessment of new technologies (Figure 1B). The framework is not intended to be a tight straitjacket that will fit all diagnostic devices. It should merely be considered an illustration of our experience with various steps that contributed to the successful accomplishing of two randomized controlled trials that we performed with PET in NSCLC. The results of these trials were implemented in guidelines, which subsequently changed the situation of patients with suspected NSCLC in our region.

Diagnostic accuracy of positron emission tomography in non-small-cell lung cancer

In NSCLC, evaluation of the extent of disease in the mediastinum is an important step in deciding about further therapy. In general, patients with preoperative identifiable N2 disease are excluded from thoracotomy. Conventional work-up of the mediastinum often includes non-invasive imaging with CT, followed by invasive exploration through mediastinoscopy. In a meta-analysis of 18 studies, including more than a thousand patients, the negative predictive value of PET was found to be 93%, which is at least as good as that of mediastinoscopy in daily clinical practice.[6] However, improved diagnostic performance does not necessarily translate into meaningful changes in clinical decisions, even at the level of this relatively simple diagnostic problem. In fact, in spite of the burgeoning number of PET studies, the 2003 guidelines on NSCLC staging[7] do not clearly recommend that mediastinoscopy can be omitted if a PET scan is negative.

Moreover, in general it will be impossible to measure impact on clinical outcomes in accuracy studies, because the design of such studies requires that the new imaging technique is assessed out of the clinical context. For example, clinically relevant PET findings (such as the suspicion of a distant metastasis in an otherwise resectable cancer) need to be confirmed because of the risk that this might be a false positive finding.[8] Since this is not acceptable in the context of an accuracy study on mediastinal staging, it needs to be proven that such confirmation is feasible in daily clinical practice.[9] Further, in clinical practice compared with accuracy studies, it is less common to have dichotomous test results,[10] and grey areas prevail with grading of diagnostic suspicion arising from a variety of test results. This contributes to the difficulty of assessing how the technique will perform in clinical practice when it is used in combination with other diagnostic interventions. How will the clinician use the newly provided diagnostic information and how will this affect the performance of other investigations and treatment?

In many ways, the level of information provided by accuracy studies can be compared to that from phase II studies in the evaluation of treatments. These studies are a prerequisite in the process of building evidence of activity, but not suited to producing information on the difference between the effects of a new drug and those of other competing drugs or devices in routine practice.

The framework

Our PET-evaluation strategy started some years before the actual installation of the PET-scanner in our region. Details of these steps in this framework are discussed below (Figure 1B).

I. ***Literature Analysis***

When searching for evidence for determining the value of a diagnostic imaging technique, typically data on diagnostic accuracy dominate the literature. Aggregation of such data should be done in a systematic review. Standards for the design, execution and reporting of accuracy studies and subsequent meta-analysis are now in place: in doing so we recommend applying the guidelines of the Cochrane Collaboration on reviewing techniques (www.cochrane.org) and initiatives such as Standards for Reporting of Diagnostic Accuracy (STARD)[5] and the Quality Assessment of Studies of Diagnostic Accuracy included in Systematic Reviews (QUADAS) instrument[11] for reporting diagnostic accuracy and for assessing the quality of accuracy studies, respectively. Without positive results from accuracy studies, preferably summarized in systematic reviews, higher-level efficacy studies are not warranted.

In 1996, a comprehensive report from the MDRC Technology Assessment Program was published with a systematic review of the literature on ^{18}FDG-PET as a diagnostic test for potential applications mainly in neurology, cardiology, and oncology.[12] Motivated by positive accuracy studies further research was suggested to define the impact of ^{18}FDG-PET on treatment decision making and on outcomes, in comparison with existing techniques.

II. ***Exploiting Clinical Data Obtained Prior to the Introduction of a New Test***

A. *Signaling (in-) efficiency and assessing potential yield in daily practice.*

To assess the potential benefit of a new test in clinical practice, it is necessary to have detailed knowledge of the situation prior to the introduction of that test. Local facilities, individual expertise, and diagnostic work-up practices may vary substantially even within a relatively small geographic area.[13] Due to such variations, potential benefit of a new device may differ between hospitals or practices. Furthermore, the observed level and nature of the (in-) efficiency provides the parameters required for sample sizes and other statistical considerations of new studies, Finally, data on the regional situation will help to interpret results from other studies and to assess the generalizability. Preferably such studies are performed on patient files and electronic registries because those reveal the actual behavior. Inclusion of different types of institutes will improve external validity of the results and has the additional benefit that a committed network of investigators is formed for further research.

We reviewed clinical practice, yield, and costs of preoperative staging for suspected NSCLC in the medical records of all patients diagnosed between 1993 and 1995 in an academic and a large community hospital.[13] Crosslinking with the Dutch Cancer Registry and the Pathological Anatomical National Registry provided complete surgical, histopathological, and follow-up data. We found a high adherence to international guidelines, despite practice variation between the two hospitals. Hospitals differed in the setting of diagnostic staging (hospitalization, outpatient setting) and the extent of mediastinoscopy use. Approximately, half of the operations for presumed resectable NSCLC proved futile. We linked this to limitations of the diagnostic tests undertaken at each level of the TNM staging process. Together with the literature survey on ^{18}FDG-PET, these data clearly indicated that there was room for improvement in the pre-operative diagnostic process by ^{18}FDG-PET.

B. *Mining Histopathological Data.*

In oncology, stored surgical specimens can also provide useful information on the potential impact of a new diagnostic test. The example hereunder illustrates this.

When several studies claimed that ^{18}FDG-PET had a high accuracy and hence could qualify for lymph node staging in breast cancer and melanoma,[14,15] we were very skeptical as these results were counter-intuitive with our experience of histopathological staging of sentinel node biopsies. We measured tumor volumes obtained in sentinel node biopsies and found that the tumor load in malignant lymph nodes was far below that which might be detected by ^{18}FDG-PET.[16,17] Subsequently, large prospective clinical studies confirmed these findings.[18-20] In our *in-vitro* melanoma study[16] we combined physical PET principles and epidemiological data to conclude that only a select group of patients might have sufficient tumor volumes to be detected by ^{18}FDG-PET. Consequently, reviewing existing datasets rather than performing (expensive) prospective clinical studies led us to conclude that PET should not be used routinely for lymph node staging in breast cancer and melanoma.

Obviously, these first steps of the framework (literature analyses and assessment of clinical data prior to the introduction of the technique) can be done simultaneously.

III. *Decision Modeling*

Decision analysis models the cost-effectiveness of a new diagnostic device. The model can combine results of clinical studies that cover different health care steps. In the presence of many alternative diagnostic strategies, decision analysis can help to identify the most promising diagnostic tests or algorithms for further research.[6,21] The data input is usually based on a meta-analysis of accuracy studies.[21] Unfortunately, accuracy studies often fail basic quality standards (e.g., independence of test interpretation, sample size, and case selection).[22,23] In addition, decision analyses studies require

a large number of assumptions to make decision problems tractable.[3] As a consequence, decision analysis is often only of limited value when faced with the complexity of daily clinical practice and the need to make decisions with respect to clinically meaningful outcomes. As more information is generated via clinical studies, fewer assumptions are required for decision modeling.[24]

> With the input of the data collected in two Dutch hospitals we considered three PET scenarios in a modeling approach: PET upfront in every patient suspected of NSCLC (1), PET after standard imaging, but prior to invasive staging (2) and PET only in patients considered operable and resectable after medical imaging and mediastinoscopy (3). From a cost perspective, the second option was considered most promising.[24]

IV. ***'Clinical-Value' Studies***

Studies that determine therapeutic plans before and after the application of a new test are sometimes referred to as 'clinical-value' studies[25] or simply 'before-after' studies.[26] By means of questionnaires, assessments of diagnostic probabilities and provisional treatment plans are made, first without the information contributed by the imaging device and then with the information.[27] In a third questionnaire the physician is asked to retrospectively grade the usefulness of this additional information in diagnostic understanding and the choice of therapy.

Credibility of such studies depends on a high quality design. Specific clinical questions should be addressed, consecutive (unselected) patients presenting with a clinical problem should be entered and changes in diagnostic certainty and therapeutic choices should be described in sufficient detail.[26] Even with attention to these issues, limitations of the clinical value study include discrepancies between the reported intention and actual clinical behavior, expectation bias, and limited generalization. The clinical value design is most useful when the availability of the new technique is still limited; in the run-up to more complex randomized studies every patient subjected to the technique can be included to provide relevant information. Standardized feed-back also helps learning from experience of both clinicians and diagnosticians. Further, in rare diseases and indications where randomized controlled trials (RCTs) are impossible, the clinical-value study may be the highest level of evidence possible.

> Since its introduction at the VUmc in 1997 the effect of every ^{18}FDG-PET scan was evaluated prospectively. Clinicians completed questionnaires just before, immediately after, and several months after the scan to study diagnostic understanding and management changes. In three years more than 600 consecutive patients were included (response 95%); half of those were referrals from outside the VUmc. Diagnostic understanding increased significantly in more than 70%, and management was changed for the benefit of the patients in 40% of all cases. The added value of the scan differed by indication. A subgroup was referred to the ^{18}FDG-PET center because of suspected NSCLC[28] with diagnostic dilemmas, such as unclear radiological findings. After PET, clinicians reported an increase in diagnostic under-

standing in 84% and beneficial management changes in 50%, mostly cancelled surgery (35%). Appreciation of ^{18}FDG-PET increased with time. Studies with similar designs in Australia[29] and the Unites States[30] also reported significant management changes (67 and 61%, respectively) due to ^{18}FDG-PET.

If a clinical-value study fails to show improved diagnostic understanding or therapeutic impact of an indication it should probably be removed from the list of potential cost-effective tests, whereas promising results warrant further investigation.

V. *Randomized Controlled Trials*

The extent to which a patient may ultimately benefit from the addition of a new imaging technique (e.g., in terms of a reduction in iatrogenic toxicity or improvement in [disease-free] survival) can only be investigated through a comparison of the full implementation of the technique added to, or in (partial) substitute of the conventional process. As only moderate benefit on patient outcome should be expected from any innovation, it is essential that both systematic and random errors are minimized. Balancing both known and unknown prognostic variables by randomly assigning patients to the new test or to the control group is the most efficient way to minimize error. Randomized controlled trials also have several qualities and benefits that arise not from the act of randomization itself but from the fact that they have many features of high-quality research[31]: a written protocol provides transparency, the pre-calculation of a sample size often allows some exploration of patient and tumor characteristics that might be outcome related, and direct cost comparisons are possible in a 'real-life' setting.

Several aspects are of particular importance for RCTs involving diagnostic imaging. In therapeutic RCTs the outcome is usually measured in terms of patient mortality and morbidity. However, diagnostic tests serve to allocate appropriate therapy to patients. A reasonable outcome measure for such studies is the extent to which appropriate therapy is applied, depending on the condition that the new test does not alter the definitions of staging, and treatment for each stage is already established. For example, there was a broad agreement among clinicians that treatment options of NSCLC patients would be clear when the diagnostic process had been completed. In addition, because ^{18}FDG-PET did not produce a new stage classification, a suitable (intermediate) outcome measure could be the reduction of iatrogenic morbidity (translated into unnecessary operations) rather than survival.[32]

For many diagnostic tests it is likely that they will first be applied as 'add-on' to conventional work-up. This is a relatively straightforward and safe approach. The moment of addition can be clearly defined in terms of logistics, as it is only necessary to have access to the technique within a reasonable time frame. However, for some tests the challenge is to study whether it can be applied earlier in the diagnostic process and may substitute other procedures. In such cases relevant endpoints could be shortening the work-up period, reducing morbidity by obviating invasive procedures, or reducing costs.

The strategy in the control group determines the extent of the contrast, and is therefore essential for the interpretation of the results. Usually the choice is between current clinical practice and 'state-of-the-art' procedures. Current clinical practice, carefully documented as the control strategy, will provide meaningful answers for the clinical community involved.[3] Our baseline study showed variances in clinical work-up between hospitals. However, through several interdisciplinary sessions in the preparation phase of the RCT, a common diagnostic workup protocol could be agreed upon.

In 1998 nine hospitals in our region enrolled 188 patients suspected for potentially resectable NSCLC. These patients were randomly allocated to the conventional work-up approach or to the same approach with ^{18}FDG-PET performed just prior to mediastinoscopy or thoracotomy.[32] The trial included about 65% of all eligible patients who were diagnosed in these institutes during that year. ^{18}FDG-PET-positive findings had to be confirmed by histology or ignored. In the conventional work-up group patients were managed as in our retrospective study. In the ^{18}FDG-PET add-on group the number of futile thoracotomies was reduced by 50%. The fact that hat objective criteria for endpoints, and clinical consensus about management of patients after diagnosis are important, is illustrated by a recently published second RCT of ^{18}FDG-PET in NSCLC.[33] In our study, 'futile thoracotomy' pertained to objective criteria (e.g., benign lung lesion, histopathological criteria, stage IIIB disease, explorative thoracotomy for any other reason or recurrent disease, or death from any cause within 1 year after randomization), whereas in the Australian trial the surgeon's decision was taken as the gold standard without validation against follow-up information (e.g., early recurrence).[34]

Our next RCT of ^{18}FDG-PET in NSCLC addressed the question of replacing conventional work-up with a ^{18}FDG-PET-scan, as had been simulated in the modeling study.[24] Between 1999 and 2001, 465 patients were enrolled by 23 hospitals. The study showed that application of ^{18}FDG-PET as the initial test had similar overall accuracy compared to traditional work-up but failed to reduce the number of tests.[35]

The result of RCTs on a diagnostic test (-sequence) should be seen in the context of patient management. In our area, there was clinical consensus that combined modality therapy (including neoadjuvant chemotherapy) should be given in case of locally advanced NSCLC. However, if an observational study on a diagnostic test identifies new prognostic subsets with unclear implication for therapy, a RCT should follow to evaluate the result of various interventions by subset rather than a trial to study the test itself.

VI. *Economic Evaluation*

One of the advantages of adding an economic evaluation to a RCT is the possibility of concurrent data collection, having the diagnostic test as the essential contrast. Health technology assessment offers a range of techniques for the evaluation of health care activities. The most common approach to economic evaluation in diagnostics is a cost-effectiveness analysis. In this type of analysis, outcome is expressed in natural units such as operations avoided or life years saved. Direct and indirect costs can be distinguished. Direct costs are defined as the resources related to the study intervention, e.g., inpatient admission, medical procedures, surgery, pharmaceutical drugs, and laboratory tests. In economic evaluations these costs are always taken into account. The importance of the other types of costs depends on the research question and the perspective (e.g., societal or payer).[36]

Several factors are specific to costing of diagnostic procedures. Diagnostic equipment is usually applied to many indications. For example ^{18}FDG-PET is also applied to other oncological and non-oncological indications. From an economic point of view the number of ^{18}FDG-PET scans performed for these indications should also be taken into account in the calculation of the cost-price of one scan. However, when the procedure is cost-effective for a certain indication one cannot automatically assume the total production capacity being filled up with this indication. In theory, one would want to tailor the required capacity of diagnostic equipment to the cost-effectiveness for different indications. Another dilemma emerges when the capacity to perform the procedure is limited. This may either result in waiting lists, or, for the sake of the RCT, priority can be given resulting in unrealistically short waiting times. In such cases indirect non-medical costs of waiting times should also be considered in the economical evaluation.

> Our RCT on add-on ^{18}FDG-PET provided direct data for comparison of costs in relation to diagnosis and therapy. Scenario analyses included various hospital settings, tracer accessibility and scenarios for ^{18}FDG-PET-scan usage.[37] All scenarios proved favorable for PET. The major cost driver was the number of hospital days related to recovery from surgery.

VII. *Before-and-After Implementation*

Studying the situation before-and-after the implementation of the procedure including ^{18}FDG-PET closes the circle of studies evaluating the cost-effectiveness of a new diagnostic device. A full appreciation of the technique should take into account all perceptions, quality, and costs of its implementation.

Data from the Regional Cancer Center Registry, where the PLUS-study was active, indicated that the results have a substantial and lasting impact. Since the guidelines were implemented in 2000, the number of lung resections dropped with an absolute 20% (corresponding to an estimated 50% reduction in unnecessary thoracotomies) compared to the average over the five years preceding to that (Figure 2).[38] A formal study is planned to investigate perceptions, quality and costs in this region of the Netherlands.

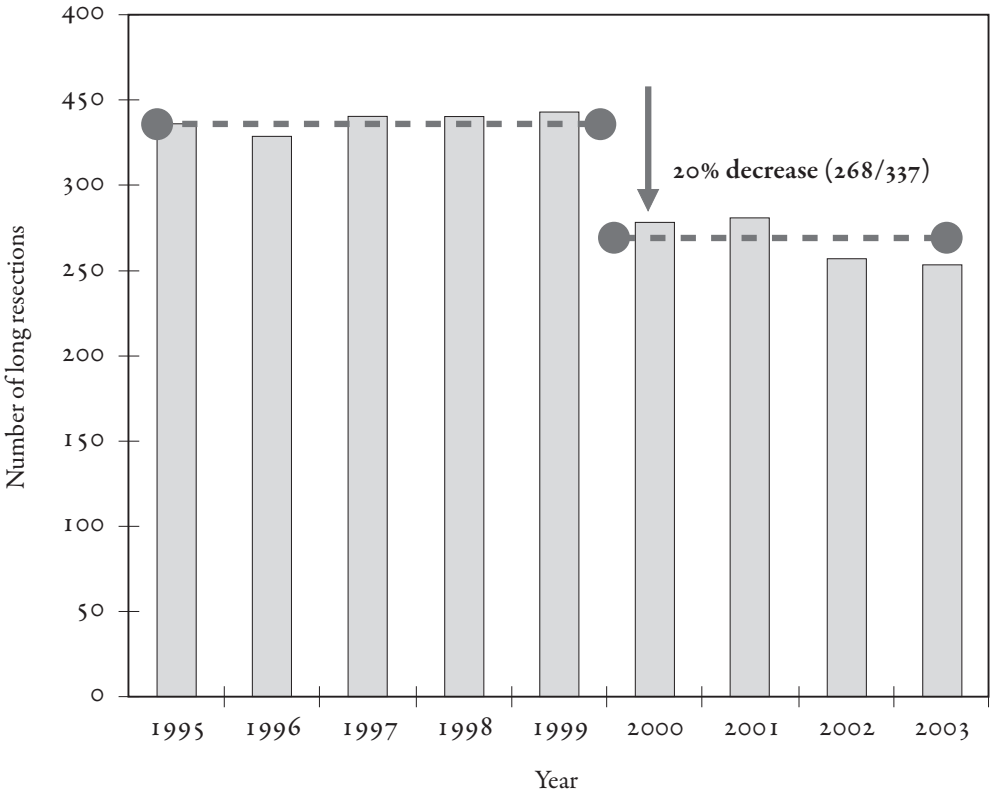


FIGURE 2. Number of lung resections before and after the implementation of a guideline on the use of Positron Emission Tomography in non-small cell lung cancer in 2000. The region of the Comprehensive Cancer Centre Amsterdam where the guideline was introduced serves 2.6 million inhabitants (Source: Netherlands Cancer Registry, Visser O. Number of lung resections in the IKA region, March 2005).

References

1. Fineberg HV. Evaluation of computed tomography: achievement and challenge. AJR *Am J Roentgenol* 1978; 131(1):1-4.
2. Steinberg EP. Magnetic resonance coronary angiography--assessing an emerging technology. *N Engl J Med* 1993; 328(12):879-880.
3. Hunink MG, Krestin GP. Study design for concurrent development, assessment, and implementation of new diagnostic imaging technology. *Radiology* 2002; 222(3):604-614.
4. Balk E, Lau J. PET scans and technology assessment: deja vu? *JAMA* 2001; 285(7):936-937.
5. Bossuyt PM, Reitsma JB. The STARD initiative. *Lancet* 2003; 361(9351):71.
6. Gould MK, Kuschner WG, Rydzak CE et al. Test performance of positron emission tomography and computed tomography for mediastinal staging in patients with non-small-cell lung cancer: a meta-analysis. *Ann Intern Med* 2003; 139(11):879-892.
7. Silvestri GA, Tanoue LT, Margolis ML, Barker J, Detterbeck F. The noninvasive staging of non-small cell lung cancer: the guidelines. *Chest* 2003; 123(1 Suppl):147S-156S.
8. Hoekstra CJ, Stroobants SG, Hoekstra OS et al. The value of [18F]fluoro-2-deoxy-D-glucose positron emission tomography in the selection of patients with stage IIIA-N2 non-small cell lung cancer for combined modality treatment. *Lung Cancer* 2003; 39(2):151-157.
9. Pieterman RM, van Putten JW, Meuzelaar JJ et al. Preoperative staging of non-small-cell lung cancer with positron-emission tomography. *N Engl J Med* 2000; 343(4):254-261.
10. Lardinois D, Weder W, Hany TF et al. Staging of non-small-cell lung cancer with integrated positron-emission tomography and computed tomography. *N Engl J Med* 2003; 348(25):2500-2507.
11. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; 3(1):25.
12. Flynn K, Adams E. Technology Assessment: Positron Emission Tomography. HTA record 988699, 1996. 1-10-1996.
13. Herder GJ, Verboom P, Smit EF et al. Practice, efficacy and cost of staging suspected non-small cell lung cancer: a retrospective study in two Dutch hospitals. *Thorax* 2002; 57(1):11-14.
14. Greco M, Crippa F, Agresti R et al. Axillary lymph node staging in breast cancer by 2-fluoro-2-deoxy-D-glucose-positron emission tomography: clinical evaluation and alternative management. *J Natl Cancer Inst* 2001; 93(8):630-635.
15. Rinne D, Baum RP, Hor G, Kaufmann R. Primary staging and follow-up of high risk melanoma patients with whole-body 18F-fluorodeoxyglucose positron emission tomography: results of a prospective study of 100 patients. *Cancer* 1998; 82(9):1664-1671.
16. Mijnhout GS, Hoekstra OS, van Lingen A et al. How morphometric analysis of metastatic load predicts the (un)usefulness of PET scanning: the case of lymph node staging in melanoma. *J Clin Pathol* 2003; 56(4):283-286.
17. Torrenga H, Licht J, van der Hoeven JJ, Hoekstra OS, Meijer S, van Diest PJ. Re: Axillary lymph node staging in breast cancer by 2-fluoro-2-deoxy-D-glucose-positron emission tomography: clinical evaluation and alternative management. J *Natl Cancer Inst* 2001; 93(21):1659-1661.
18. Wagner JD, Schauwecker D, Davidson D et al. Prospective study of fluorodeoxyglucose-positron emission tomography imaging of lymph node basins in melanoma patients undergoing sentinel node biopsy. *J Clin Oncol* 1999; 17(5):1508-1515.
19. Wagner JD, Schauwecker DS, Davidson D, Wenck S, Jung SH, Hutchins G. FDG-PET sensitivity for melanoma lymph node metastases is dependent on tumor volume. *J Surg Oncol* 2001; 77(4):237-242.
20. Wahl RL, Siegel BA, Coleman RE, Gatsonis CG. Prospective multicenter study of axillary nodal staging by positron emission tomography in breast cancer: a report of the staging breast cancer with PET Study Group. *J Clin Oncol* 2004; 22(2):277-285.
21. Gould MK, Sanders GD, Barnett PG et al. Cost-effectiveness of alternative management strategies for patients with solitary pulmonary nodules. *Ann Intern Med 2003*; 138(9):724-735.

22. Gould MK, Maclean CC, Kuschner WG, Rydzak CE, Owens DK. Accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions: a meta-analysis. *JAMA* 2001; 285(7):914-924.
23. Toloza EM, Harpole L, McCrory DC. Noninvasive staging of non-small cell lung cancer: a review of the current evidence. *Chest* 2003; 123(1 Suppl):137S-146S.
24. Verboom P, Herder GJ, Hoekstra OS et al. Staging of non-small-cell lung cancer and application of FDG-PET. A cost modeling approach. *Int J Technol Assess* Health Care 2002; 18(3):576-585.
25. Freedman LS. Evaluating and comparing imaging techniques: a review and classification of study designs. *Br J Radiol* 1987; 60(719):1071-1081.
26. Guyatt GH, Tugwell PX, Feeny DH, Drummond MF, Haynes RB. The role of before-after studies of therapeutic impact in the evaluation of diagnostic technologies. *J Chronic Dis* 1986; 39(4):295-304.
27. Wittenberg J, Fineberg HV, Ferrucci JT, Jr. et al. Clinical efficacy of computed body tomography, II. *AJR Am J Roentgenol* 1980; 134(6):1111-1120.
28. Herder GJ, van Tinteren H, Comans EF et al. Prospective use of serial questionnaires to evaluate the therapeutic efficacy of 18F-fluorodeoxyglucose (FDG) positron emission tomography (PET) in suspected lung cancer. *Thorax* 2003; 58(1):47-51.
29. Kalff V, Hicks RJ, MacManus MP et al. Clinical impact of (18)F fluorodeoxyglucose positron emission tomography in patients with non-small-cell lung cancer: a prospective study. *J Clin Oncol* 2001; 19(1):111-118.
30. Hillner BE, Tunuguntla R, Fratkin M. Clinical decisions associated with positron emission tomography in a prospective cohort of patients with suspected or known cancer at one United States center. *J Clin Oncol* 2004; 22(20):4147-4156.
31. Abel U, Koch A. The role of randomization in clinical studies: myths and beliefs. *J Clin Epidemiol* 1999; 52(6):487-497.
32. van Tinteren H, Hoekstra OS, Smit EF et al. Effectiveness of positron emission tomography in the preoperative assessment of patients with suspected non-small-cell lung cancer: the PLUS multicentre randomised trial. *Lancet* 2002; 359(9315):1388-1393.
33. Viney RC, Boyer MJ, King MT et al. Randomized controlled trial of the role of positron emission tomography in the management of stage I and II non-small-cell lung cancer. *J Clin Oncol* 2004; 22(12):2357-2362.
34. van Tinteren H, Smit EF, Hoekstra OS. FDG-PET in addition to conventional work-up in non-small-cell lung cancer. *J Clin Oncol* 2005; 23(7):1591-1592.
35. Herder GJ. Traditional versus up-front 18FDG PET staging of non-small cell lung cancer (NSCLC): A Dutch Co-operative randomized study. *J Clin Oncol* 2004; 22(14S):7000.
36. Drummond MF, Richardson WS, O'Brien BJ, Levine M, Heyland D. Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1997; 277(19):1552-1557.
37. Verboom P, van Tinteren H, Hoekstra OS et al. Cost-effectiveness of FDG-PET in staging non-small cell lung cancer: the PLUS study. *Eur J Nucl Med Mol* Imaging 2003; 30(11):1444-1449.
38. Visser O. Lung resections in the region of the Comprehensive Cancer Center Amsterdam, The Netherlands. personal communication . 2004.
39. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991; 11(2):88-94.

CHAPTER 9

Summary and Epilogue

Summary and Epilogue

Positron Emission Tomography (PET) is a relatively new nuclear medicine imaging technique that allows the visualization of biochemical processes in tissues. Therefore, PET has the ability to complement traditional imaging modes such as computed tomography (CT) and magnetic resonance imaging (MRI), which provide information on anatomical structures. When, in 1996, a dedicated positron emission tomography (PET)-scanner was planned in the VU University Medical Centre, it was decided to focus the clinical PET research to the identification of cost-effective applications in routine cancer practice. This thesis provides the theoretical back-bone of this process and describes its application in non-small cell lung cancer.

In **chapter 1** an introduction is given to the complexities of diagnosing and staging of disease in general and of non-small-cell lung cancer (NSCLC) in particular. The actual situation of management of NSCLC had just been studied in two hospitals in the geographical region of interest and showed substantial residual shortcomings.[1] It was found that in nearly 50% of operated patients surgical intervention failed because of irresectable tumor, benign lesions at surgery or because the disease recurred or metastasized within one year. At the same time, PET with ^{18}F-fluorodexoyglucose (FDG-PET) had shown promising results in accuracy studies in several aspects of the NSCLC staging.[2] However, improved accuracy does not necessarily imply clinical usefulness, e.g. better patient management and improved clinical outcome.[3] The assessment of the value of a diagnostic test to patient management best follows a multi-phase hierarchical process, as has been described by Fryback and others.[4] The process encompasses the technical aspects of the test such as image quality and reproducibility, diagnostic accuracy, diagnostic and therapeutic impact, patient outcomes and finally, the cost-benefit analysis of the introduction of the technology. Obviously, demonstration of efficacy at each lower level is logically necessary, but not sufficient, to assure efficacy at a higher level. Given the promising data from accuracy studies and prompted by the result of modeling studies [5] we considered that evaluation of the added value of FDG-PET should preferably be investigated in a direct comparison of the conventional strategy with a strategy that included the new device as 'add-on'. To minimize potential bias in that comparison, due to measurable and non-measurable patient- and tumor characteristics and diagnostician/clinician decisions, we started to plan a randomized controlled trial (RCT).

To explore and improve our understanding of the potential impact of FDG-PET on management decisions in the preoperative setting of NSCLC, we first designed a 'clinical value' or 'before-after' study (**chapter 2**). Through prospective use of serial questionnaires, physicians were asked to indicate their diagnostic understanding and planning for subsequent diagnostic or therapeutic steps, just before, immediately after and several months after learning the results of the PET-scan. Patients with suspected NSCLC were referred to the PET-center because non-invasive tests had failed to solve a diagnostic problem, usually unclear radiological findings. Based on 164 patients, the study revealed that PET had a positive influence on diagnostic understanding in 84% of cases and, according to the physicians, PET resulted in a beneficial change of treatment in 50%.[6] Cancelled surgeries were the

most frequently reported changes in treatment after PET (35%). Overall PET proved to be the key diagnostic tool in one of every four patients referred to PET.

For the purpose of designing the diagnostic RCT we performed a literature search (in 1996, which was reiterated in 2005) to find other randomized diagnostic imaging studies. In **chapter 3** the results of this investigation are presented. We designed a taxonomy of diagnostic RCTs (D-RCTs) accounting for main study contrasts and outcome measures. Applying this taxonomy, we were able to investigate possible time trends with respect to D-RCTs addressing patient outcomes as a function of imaging tests. Our first search included three distinctive years of publication between 1990 and 2002 at intervals of 6 years to allow capturing possible developments in time. In a second search we focused on D-RCTs performed between 1990 and 2005, in which Magnetic Resonance Imaging (MRI) played a major role in the comparison. MRI is the most recent tomographic imaging device before PET, and we expected similarity of the respective assessment processes. In our non-specific search for D-RCTs, the majority of identified studies (82%) pertained to testing different tracers or acquisition parameters with one particular device. An increase over time was observed in studies comparing devices vs. studies testing tracers. In the end, we could identify only 15 D-RCTs in which patient outcomes were studied. The disease indications varied from cardiovascular and musculoskeletal problems to cancer. In the second search we identified 13 D-RCTs studying the value of MRI in the past 15 years. Four of those were published in 2005 and 4 others between 2000 and 2005, suggesting an increased interest in D-RCTs. Three of the 13 studies applied two contrasting tests or strategies in the same patients. In the other 10 studies, patients were randomly allocated to a strategy with or without MRI. Based on these results, we concluded that, although generally advocated and appreciated, randomized controlled trials in diagnostic imaging are not frequently performed. As a consequence, evidence on (cost)-effective use of imaging tests is lacking in general.

Because our literature search revealed that D-RCTs are relatively rare and certainly non-existent in the field of PET research, we considered it useful to publish and communicate about the design and its logistics (**Chapter 4**).[7] The study was named the 'PLUS study' (PET in LUng cancer Staging). Patients were eligible if they had suspected or proven NSCLC, considered medically operable and potentially resectable by the local pulmonary physician on the basis of clinical staging procedures (i.e. clinical stage I-III). Patients were randomly allocated, just before invasive verification or therapy was considered, to either 'continue as usual' or to undergo a PET-scan. Of particular interest was the choice of outcome: we defined the primary endpoint as the difference in the number of futile thoracotomies between both groups. Thoracotomy was considered futile in case of benign lesions, pathologically proven mediastinal lymph node involvement (stage IIIA-N2), stage IIIB, explorative thoracotomy for any reason or recurrent disease or death of any cause within one year after randomization. The protocol required confirmation of clinically decisive PET results. Next to secondary endpoints, such as morbidity and duration of diagnostic and therapeutic processes, a separate costs analysis was planned. The study was powered to detect an absolute reduction of futile thoracotomies of 25% (power of 90%, alpha of 0.05 two-sided), assuming a 45% futile thoracotomy rate in the current setting as documented by our retrospective cohort analysis. Analyses would be performed according to the intention-to-diagnose principle.

In one year, 188 patients were included from nine hospitals.[8] Data from the Dutch Cancer Registry suggested that the study included 65% of all eligible patients. 96 patients were randomized to the conventional group (CWU) and 92 to the CWU + PET group (**Chapter 5**). In each group, 70% of patients had clinical stage I/II disease. A significant greater proportion of patients underwent futile thoracotomy in the CWU group than in the CWU + PET group. The relative reduction was 51% ($p = 0.003$) and the absolute difference 20% (41% futile thoracotomies in the CWU group). This result can be interpreted as five patients who needed PET to avoid one futile thoracotomy. PET correctly suggested that surgery was justified in 81% of scans, versus 71% in which PET suggested surgery was futile. The main effect of PET was to upstage patients.

Together with clinical data, data on cost items were collected. Cost items included invasive and non-invasive tests (including PET), thoracotomies and hospital days on an individual and hospital level. The costs of PET encompassed the personnel costs, the depreciation and maintenance costs, the material (tracer) costs and overhead. The cost price of PET varied between € 736 and € 1,588 depending on the hospital setting and the FDG availability.[9] The average costs per patient in the PLUS-CWU group were €9,573 and in the CWU+PET group €8,284. The major cost driver proved to be the number of hospital days related to recovery from surgery. A sensitivity analysis, varying the efficacy of PET showed that the cost results are robust in favor of PET (**Chapter 6**). In addition to the straightforward cost analysis within the PLUS-setting, three scenarios were considered by varying number of PET scans performed per day and varying hospital settings. An 'expensive' scenario considered PET in a university hospital, with on-site tracer production and with clinical and research functions. In a 'cheap' variant, 12 clinical PET-scans were made in a community hospital with FDG production and transport from elsewhere. The 'in-between' option represented a large community hospital with on-site FDG production, 8 PET-scans per day and limited research functions. The scenario analyses, using the point-estimate of efficacy of PET but varying the setting also proved to be favorable for PET, even in the expensive setting. Only when the worst efficacy (1 futile thoracotomy prevented for every 14 PET-scans) was applied to the expensive variant (university setting), the results were in favor of the CWU approach (namely €542).

Not every expert endorses the need for diagnostic randomized controlled trials.[10] Moreover, unnecessary use of resources and ethical considerations are used to argue against D-RCTs. This prompted us to explicate our thoughts about the value of randomization in diagnostic research (**Chapter 7**).[11] Briefly put, the extent to which a patient may ultimately benefit from the addition of a new diagnostic imaging technique (in terms of reduction of iatrogenic toxicity or improvement in survival) can only be investigated in a concurrent comparison of conventional strategies in a routine clinical setting without having to make unrealistic or unverifiable assumptions. Balancing both known and unknown prognostic variables by randomly assigning patients to the new and the conventional strategy is the most efficient way to do this, along with making the study as large as possible. As well as minimizing imbalances in prognostic factors, randomized studies also have several qualities and benefits that arise not from the act of random allocation itself, but from the fact that they have many

features of high-quality research, such as transparency on the base of a written protocol and a pre-calculated sample size, to mention a few.
The two randomized studies with FDG-PET in NSCLC that we performed show that pragmatic trials are feasible, although the window of opportunity may be limited. The PLUS study accrued 188 patients with clinical stage I-III in one year. The POORT study, in which PET was tested at the very first suspicion of lung cancer to substitute other tests in number, enrolled 465 patients 2 years.[12] The fact that PET-capacity in the Netherlands was limited at that time may have contributed to that success.
In **Chapter 8** we comprise our experience on the evaluation of cost-effectiveness of FDG-PET for different indications, and present a framework which basically follows the postulated hierarchical approach for the assessment of new technologies as referred to earlier. The starting point of our studies was a notion of residual inefficiency which might be amenable with the suggested improved accuracy of FDG-PET (in NSCLC). Retrospective cohort analysis substantiated this notion, specified the nature of the errors and it proved to be the main incentive for clinicians to participate in further research. The run-in experiment on the clinical value of PET was the learning curve of diagnosticians and clinicians. In the end, the coherence of this approach contributed to a very fruitful environment for collaboration and has facilitated the successful completion of the two RCTs and the appearance and implementation of working guidelines in our region.

The present

Anno 2006, the PET-situation in the Netherlands has changed substantially. The PET-scanner capacity has increased dramatically and currently most hospitals have access to fixed or mobile PET-scanners. In the Amsterdam region at least five scanners are available in a radius of about 20 kilometers. From 2004 evidence based guidelines for diagnosis and management in NSCLC have incorporated PET on a national level.[13] Ongoing studies are performed to update and refine these guidelines. It is estimated that in about 80% of the patients in the IKA region (2.6 million inhabitants, 1300 NSCLC annually) with suspected operable NSCLC a PET-scan is integrated in their work-up. A recent analysis on data from the region of the comprehensive Cancer Center of Amsterdam suggested an absolute 20% decrease in the number of thoracotomies (corresponding to an estimated 50% reduction in unnecessary thoracotomies) compared to the average over the five years preceding to that.[14]
Since the advent of the PET-scanner, improvements in the concept of the scanner and other technologies have seen the light. After 15 years of whole body PET, the integrated or hybrid PET-CT scanner has been introduced.[15-17] Several studies have claimed that the 'hardware' fused whole body anatomical (CT) and functional (PET) images have superior accuracy compared to software fusion or visual fusion (side-by-side reading, as employed in the PLUS study). A quick search in MEDLINE revealed five studies investigating the accuracy of the integrated PET/CT as compared to

the other devices in NSCLC staging [18-22]. All five studies suggested an improved accuracy for PET/CT compared to side-by side reading of PET and CT or PET alone and two of them also claimed changes in patient management.[22,21] However, the primary question should be what the size and the nature of any residual inefficiency in current NSCLC staging are (with PET integrated). Even though we are confident that adding the PET to standard diagnostic procedures reduces the number of thoracotomies, we know from the PLUS study that 20% of the thoracotomies is still unnecessary. To assess whether the integrated PET-CT could have prevented these thoracotomies we looked at the individual cases: In 9 patients of the PLUS study, the thoracotomy itself proved to be futile and in another 10 patients the disease recurred or the patients died early (within one year). Of the former 9 patients, 2 had benign disease, six were upstaged and one patient was not radically operated because the residual lung capacity precluded the pneumonectomy. In 3 of these 6 upstaged patients a mediastinoscopy had not been able to confirm the mediastinal lymph node involvement as suggested by PET. In the 10 patients in which the surgery was considered futile in the follow-up-year, 4 patients relapsed after apparent curative surgery, two with bone metastases and one with metastases in brain and skeleton after refusal of a PET. In the other patient pulmonary metastases of melanoma (primary site unknown) had been resected but disseminated involvement became apparent during follow-up. Five patients died of surgery related causes and one patient died of unknown cause. To conclude, more sophisticated non-invasive staging could have prevented at the most 3 of the 19 futile cases on the totality of 92 patients seen with PET. We recognize that these numbers are small and do not allow firm conclusions. However, they might give some indication of the expected yield of PET-CT in the context of preventing unnecessary surgery. Compared to PET alone, PET-CT clearly adds to the specificity of PET readings, and as a clinical spin-off confirmatory biopsy procedures might be conducted more efficiently. In the PLUS study, lower cervical lymph nodes positive at PET proved to be a major challenge for radiologists in community hospitals. However, such failed confirmative procedure did not affect the number of futile thoracotomies since lymph node involvement could be confirmed otherwise (albeit more invasively). With respect to the impact of PET-CT on sensitivity: we have argued that test results are not dichotomous, and this is why PET-CT, even though it consists of the same PET and CT scanning technologies as in the stand-alone situations, might help to flip the coin towards higher levels of suspicion. However, if the suggestion raised by our PLUS study evaluation of residual problems after PET is correct, we do not expect major incremental benefits on that level. Claims about improved staging at the level of the primary tumor extension[20] obviously need confirmation given the limited resolution of PET.[23] Perhaps alternative techniques to explore the mediastinum preoperatively are more promising. Endoscopic esophageal and endobronchial ultrasound-guided fine needle aspiration (EUS- and EBUS-FNA) are novel minimally invasive techniques with potential for the analysis of mediastinal lymph nodes partially complementary to mediastinoscopy and of tumor invasion in centrally located tumors.[24]

Alternatively, in case of PET-CT, logistic factors may prevail that justify a switch to this modality, even when increments of clinical effectiveness or lower costs are unlikely or difficult to prove. The concept of staging a patient with a single scan ('one-stop-shop') rather than with a battery of tests conducted over a period of time is highly attractive from both a patient as well as a management per-

spective. Moreover the perspective of reducing the PET-scan time per patient should be appealing for those currently confronted with waiting lists.
At the end of the day, whichever methodological concerns we might have about the evidence of superiority of PET-CT vs. PET, the industrial development is such that PET-CT will be the standard of practice soon, simply because whole body PET scanners are no longer being sold. An unsolved issue is whether PET-CT should be implemented upfront in the diagnostic process or just prior to mediastinal evaluation: it is not economical to perform PET-CT on every patient if CT alone would eliminate a considerable number of surgical candidates by showing disseminated disease or benign primary pulmonary lesion. Decision analysis preferably based on collected information of actual costs and scenario analyses can help to clarify this issue.

So far, we have discussed FDG-PET in the context of surgical decision-making purely in terms of TNM staging issues. However, metabolic information obtained by FDG PET also adds prognostic information at the biological level within clinical stages.[25] high uptake in tumors is prognostically unfavorable compared to lower uptake. Even though lack of standardized PET procedures impairs meta-analysis of individual studies again, the point seems to be made. How this information may be combined with other prognostic markers to develop strategies to improve the relatively poor outcome of patients with resectable lung cancer remains to be shown, but it is likely that systemic therapy will be required in subsets of patients. A similar prognostic feature of PET might be relevant in patients with locally advanced disease who are treated with combined modality therapy. Here, the role of surgery (besides that of chemoradiation therapy) is at stake, and the issue is to select the subset of patients who will benefit from surgery. We and others have shown in observational studies that metabolic behavior provided by FDG PET appears to have added prognostic value already early during systemic therapy.[26,27]
To evaluate the clinical relevance of such metabolic patterns, randomized controlled trials are required. The issue of these trials will now go beyond the impact of a diagnostic test on TNM-stage related outcomes such as futile surgery towards that of the diagnosis-intervention combination to improve survival.

The following sentences may summarize this thesis: to actually evaluate (cost-) effectiveness of diagnostic devices, studies are needed that focus on clinical relevant endpoints beyond diagnostic accuracy. It is argued and shown that in particular randomized controlled trials can produce results that are convincing and directly applicable in clinical practice. The framework of studies presented here may support the evaluation of other diagnostic devices.

References

1. Herder, GJ, Verboom, P., Smit, E. F. et al. Practice, efficacy and cost of staging suspected non-small cell lung cancer: a retrospective study in two Dutch hospitals. Thorax, 2002; 57: 11-14.
2. Flynn, K and Adams, E. Technology Assessment: Positron Emission Tomography. Original Report, 1996. 1996;
3. Freedman, LS. Evaluating and comparing imaging techniques: a review and classification of study designs. Br J Radiol, 1987; 60: 1071-1081.
4. Fryback, DG and Thornbury, J. R. The efficacy of diagnostic imaging. Med Decis Making, 1991; 11: 88-94.
5. Verboom, P, Herder, G. J., Hoekstra, O. S. et al. Staging of non-small-cell lung cancer and application of FDG-PET. A cost modeling approach. Int J Technol Assess Health Care, 2002; 18: 576-585.
6. Herder, GJ, van Tinteren, H., Comans, E. F. et al. Prospective use of serial questionnaires to evaluate the therapeutic efficacy of 18F-fluorodeoxyglucose (FDG) positron emission tomography (PET) in suspected lung cancer. Thorax, 2003; 58: 47-51.
7. van Tinteren, H, Hoekstra, O. S., Smit, E. F. et al. Toward less futile surgery in non-small cell lung cancer? A randomized clinical trial to evaluate the cost-effectiveness of positron emission tomography. Control Clin Trials, 2001; 22: 89-98.
8. van Tinteren, H, Hoekstra, O. S., Smit, E. F. et al. Effectiveness of positron emission tomography in the preoperative assessment of patients with suspected non-small-cell lung cancer: the PLUS multicentre randomised trial. Lancet, 2002; 359: 1388-1393.
9. Verboom, P, van Tinteren, H., Hoekstra, O. S. et al. Cost-effectiveness of FDG-PET in staging non-small cell lung cancer: the PLUS study. Eur J Nucl Med Mol Imaging, 2003; 30: 1444-1449.
10. Valk, PE. Do we need randomised trials to evaluate diagnostic procedures? Against. Eur J Nucl Med Mol Imaging, 2004; 31: 132-135.
11. van Tinteren, H, Hoekstra, O. S., and Boers, M. Do we need randomised trials to evaluate diagnostic procedures? For. Eur J Nucl Med Mol Imaging, 2004; 31: 129-131.
12. Herder, GJ, Kramer, H., Hoekstra, O. S. et al. Traditional Versus Up-Front [18F] Fluorodeoxyglucose-Positron Emission Tomography Staging of Non-Small-Cell Lung Cancer: A Dutch Cooperative Randomized Study. J Clin Oncol, 2006; 24: 1800-1806.
13. van Meerbeeck, JP, Koning, C. C., Tjan-Heijnen, V. C. et al. [Guideline on 'non-small cell lung carcinoma; staging and treatment']. Ned Tijdschr Geneeskd, 2005; 149: 72-77.
14. van Tinteren, H, Smit, E. F., and Hoekstra, O. S. FDG-PET in addition to conventional work-up in non-small-cell lung cancer. J Clin Oncol, 2005; 23: 1591-1592.
15. Beyer, T, Townsend, D. W., Brun, T. et al. A combined PET/CT scanner for clinical oncology. J Nucl Med, 2000; 41: 1369-1379.
16. Vogel, WV, Oyen, W. J., Barentsz, J. O. et al. PET/CT: panacea, redundancy, or something in between? J Nucl Med, 2004; 45 Suppl 1:15S-24S.: 15S-24S.
17. Townsend, DW, Carney, J. P., Yap, J. T. et al. PET/CT today and tomorrow. J Nucl Med, 2004; 45 Suppl 1:4S-14S.: 4S-14S.
18. Halpern, BS, Schiepers, C., Weber, W. A. et al. Presurgical staging of non-small cell lung cancer: positron emission tomography, integrated positron emission tomography/CT, and software image fusion. Chest, 2005; 128: 2289-2297.
19. Cerfolio, RJ, Ojha, B., Bryant, A. S. et al. The accuracy of integrated PET-CT compared with dedicated pet alone for the staging of patients with nonsmall cell lung cancer. Ann Thorac Surg, 2004; 78: 1017-1023.
20. Lardinois, D, Weder, W., Hany, T. F. et al. Staging of non-small-cell lung cancer with integrated positron-emission tomography and computed tomography. N Engl J Med, 2003; 348: 2500-2507.
21. Antoch, G, Stattaus, J., Nemat, A. T. et al. Non-small cell lung cancer: dual-modality PET/CT in preoperative staging. Radiology, 2003; 229: 526-533.
22. Keidar, Z, Haim, N., Guralnik, L. et al. PET/CT using 18F-FDG in suspected lung cancer recurrence: diagnostic value and impact on patient management. J Nucl Med, 2004; 45: 1640-1646.

23. Allen-Auerbach, M, Yeom, K., Park, J. et al. Standard PET/CT of the Chest During Shallow Breathing Is Inadequate for Comprehensive Staging of Lung Cancer. J Nucl Med, 2006; 47: 298-301.
24. Annema, JT, Hoekstra, O. S., Smit, E. F. et al. Towards a minimally invasive staging strategy in NSCLC: analysis of PET positive mediastinal lesions by EUS-FNA. Lung Cancer, 2004; 44: 53-60.
25. Vansteenkiste, J, Fischer, B. M., Dooms, C. et al. Positron-emission tomography in prognostic and therapeutic assessment of lung cancer: systematic review. Lancet Oncol, 2004; 5: 531-540.
26. Hoekstra, CJ, Stroobants, S. G., Smit, E. F. et al. Prognostic relevance of response evaluation using [18F]-2-fluoro-2-deoxy-D-glucose positron emission tomography in patients with locally advanced non-small-cell lung cancer. J Clin Oncol, 2005; 23: 8362-8370.
27. Cerfolio, RJ, Bryant, A. S., Winokur, T. S. et al. Repeat FDG-PET after neoadjuvant therapy is a predictor of pathologic response in patients with non-small cell lung cancer. Ann Thorac Surg, 2004; 78: 1903-1909.

Samenvatting en Epiloog

Samenvatting en Epiloog

Positron Emissie Tomografie (PET) is een betrekkelijk nieuwe nucleair geneeskundige beeldvormende techniek die het mogelijk maakt om biochemische processen in weefsels zichtbaar te maken. Om die reden kan PET een complementaire rol spelen bij traditionele beeldvormende technieken zoals 'computed tomography' (CT) en 'magnetic resonance imaging' (MRI) die anatomische informatie opleveren.

Toen het VU Medisch Centrum in 1996 overging tot de aanschaf van een PET-scanner werd besloten de klinische wetenschappelijk aandacht te richten op het identificeren van kosteneffectieve toepassingen in de klinische praktijk. Dit proefschrift biedt de theoretische ondersteuning voor dit proces en beschrijft haar toepassing bij niet-klein-cellige longkanker.

In **hoofdstuk 1** wordt ingegaan op de complexiteit van diagnosticeren en stadiëren in het algemeen en bij niet-klein-cellige longkanker (NSCLC) in het bijzonder. In een recent cohort onderzoek, uitgevoerd in twee ziekenhuizen in de relevante geografische regio, werden tekortkomingen gevonden in de routinematige diagnostiek bij longkanker.[1] In bijna 50% van de geopereerde patiënten bleek de chirurgische interventie niet geslaagd vanwege niet-verwijderbaar tumorweefsel, goedaardige aandoeningen of omdat de ziekte binnen een jaar terug kwam of de patiënt overleed in die periode. In dezelfde periode lieten elders studies naar PET met ^{18}F-fluorodexoyglucose als 'tracer' (FDG-PET), veelbelovende resultaten zien wat betreft de accuratesse voor verschillende aspecten van NSCLC-stadiëring.[2] Echter, betere accuratesse betekent niet automatisch dat het klinisch bruikbaar is, dat wil zeggen dat het leidt tot betere diagnostiek en behandeling van patiënten en uiteindelijk tot betere uitkomsten.[3]

Het evalueren van de waarde van een diagnostische test kan over het algemeen het best geschieden volgens een gefaseerd hiërarchisch proces, zoals in de literatuur beschreven door Fryback en anderen.[4] Een dergelijk proces omvat een vijftal niveaus: technische aspecten, zoals kwaliteit van de beelden en reproduceerbaarheid – meerwaarde op het diagnostisch inzicht –, invloed op behandelingsbeslissingen, uitkomsten voor patiënten en als hoogste niveau geldt een kosten-batenanalyse bij introductie van de nieuwe test. Het aantonen van effectiviteit op een bepaald niveau is een voorwaarde, maar is op zichzelf geen garantie, voor effectiviteit op een hoger niveau.

Gegeven de veelbelovende resultaten van de accuratesse studies en geleid door het resultaat van een modelmatige studie naar de kosteneffectieve toepassing van de nieuwe test,[5] waren wij van mening dat de toegevoegde waarde van FDG-PET het best kon worden bestudeerd door gelijktijdig het huidige diagnostische beleid te vergelijken met hetzelfde beleid, maar waar dan FDG-PET aan toegevoegd zou zijn. Om mogelijke verstoringen door verschillen in patiënt- of tumorkarakteristieken zo klein mogelijk te maken leek een gerandomiseerde studie (RCT) de meest geschikte studieopzet (zoals ook gebruikelijk bij studies naar nieuwe geneesmiddelen).

Om het begrip ten aanzien van de potentiële invloed van PET op diagnostisch handelen en op behandelingsbeleid in de pre-operatieve situatie van NSCLC te vergroten hebben een we eerst een 'clinical value' of 'voor – na' studie uitgevoerd (**hoofdstuk 2**).[6] Met behulp van vragenlijsten werden

artsen – gaande het diagnostische proces – verzocht aan te geven wat hun diagnostische begrip was en welke behandelingsstrategie ze zouden volgen net voor, meteen na en 3 maanden na het verkrijgen van de resultaten van de PET-scan. Patiënten, verdacht van NSCLC, werden naar het PET centrum verwezen indien niet-invasieve testen een bepaald diagnostisch probleem, meestal van radiologische aard, niet hadden kunnen oplossen. Op basis van 164 patiënten liet de studie zien dat PET een gunstige invloed had op het diagnostische begrip in 84% van de patiënten en dat, volgens de clinici, PET het behandelingsbeleid in 50% in positieve zin had bepaald. Afstel van chirurgie op basis van PET werd het meest frequent gerapporteerd. PET bleek een sleutelrol te hebben gespeeld bij 1 op de 4 patiënten in de studie.

Ten behoeve van het ontwerp van de diagnostische RCT voerden we een literatuuranalyse uit (in 1996 en herhaald in 2005) om andere diagnostische RCTs (D-RCTs) te vinden. In **hoofdstuk 3** worden de resultaten van die zoekactie gepresenteerd. We ontwikkelden een taxonomie van diagnostische RCTs op basis van de vergelijking binnen studies en de uitkomstmaten. De taxonomie maakte het mogelijke om eventuele veranderingen in de tijd, in uitkomstmaten in relatie tot de beeldvormende testen, te onderzoeken. De eerste zoekactie omvatte 3 verschillende jaren over een tijdspanne van 1990 tot 2002. De tweede zoekactie was volledig gericht op studies waarbij MRI als diagnostisch instrument onderwerp van studie was, en omvatte in de gehele periode van 1990 tot en met 2005. MRI kan worden beschouwd als de meest recente tomografische beeldvormende techniek vòòr de komst van PET en we verwachtten overeenkomsten in de evaluatiemethoden. Uit de aspecifieke zoekactie naar D-RCTs bleek dat het grootste deel van de gevonden studies betrekking had op het vergelijken van 'tracers' of technische uitvoeringskenmerken behorende bij een enkel diagnostische instrument. Uiteindelijk konden we slechts 15 D-RCTs vinden waarin patiëntuitkomsten werden bestudeerd. De ziekte-indicaties varieerden van cardiovasculaire en gewricht- en spierproblemen tot kanker. In de tweede zoekactie, die in totaal 16 jaar omvatte, vonden we 13 unieke RCTs waarin de waarde van MRI werd onderzocht. Vier van deze studies werden gepubliceerd in 2005 en 4 in de periode van 2000 tot 2005. In 3 van de 13 studies werden twee onderzoekstechnieken in dezelfde patiënt toegepast. In de overige 10 studies werden de patiënten willekeurig verdeeld over de beide strategieën. Op basis van de resultaten concluderen we onder andere dat, hoewel in veel literatuur aangehaald als optimale methode voor de evaluatie van diagnostische technieken, RCTs niet met regelmaat worden uitgevoerd. Daardoor ontbreekt in het algemeen het bewijs voor een (kosten-)effectieve toepassing van beeldvormende technieken.

Omdat het literatuuronderzoek liet zien dat D-RCTs schaars zijn en zeker op het gebied van PET nog niet was uitgevoerd, achtten we het van belang om het ontwerp en de logistiek van de studie te publiceren (**hoofdstuk 4**).[7] De studie kreeg als acroniem de naam PLUS mee (PET in LUng cancer Staging). Om aan de studie deel te nemen moesten de patiënten aan een aantal criteria voldoen: zo moest onder andere sprake zijn van een verdenking op, of bewezen NSCLC op basis van klinische stadiering (klinisch stadium I-III) en de patiënten moesten medisch operabel en potentieel resectabel zijn naar het oordeel van de plaatselijk behandelend longarts. Net voordat een invasieve ingreep (diagnostisch of therapeutisch) aan de orde kwam werden de patiënten willekeurig ingedeeld in de groep 'conventionele strategie' of 'conventionele strategie met PET-scan'. Veel aandacht werd

besteed aan de definitie van het eindpunt. Als primair eindpunt werd gekozen voor het verschil in aantal 'onnodige of onterechte' chirurgische ingrepen (thoracotomieën) tussen beide groepen. Een thoracotomie werd onnodig of onterecht geacht indien er sprake was van: goedaardige aandoeningen, pathologisch bewezen mediastinale lymfeklieren (stadium IIIA-N_2), stadium IIIB of hoger, exploratieve chirurgie of terugkerende ziekte of sterfte binnen één jaar na randomisatie. Het protocol vereiste pathologische bevestiging van beleidsbepalende PET-positieve verdachte laesies. Naast secundaire eindpunten als morbiditeit en duur van het diagnostische proces werd een kostenanalyse gepland. De studie werd zodanig ontworpen dat met 95% zekerheid een absoluut verschil van 25% in het aantal thoracotomieën zou kunnen worden vastgesteld (met onderscheidend vermogen van 90%), uitgaande van 45% onterechte thoracotomieën in de huidige situatie (zoals uit het vooronderzoek was gebleken). Analyses zouden worden uitgevoerd volgens het 'intention-to-diagnose' principe (patiënten worden in de analyses meegenomen bij de groep waar ze volgens randomisatie aan toe behoren).

In één enkel jaar werden 188 patiënten uit 9 verschillende ziekenhuizen in de studie opgenomen.[8] Gegevens uit de kankerregistratie van de IKA-regio suggereren dat dit aantal overeenkomt met 65% van de potentieel beschikbare populatie in dat jaar. 96 patiënten werden ingedeeld in de 'conventionele strategie'-groep (CWU) en 92 in de 'conventionele + PET'-groep (**hoofdstuk 5**). In beide groepen had 70% van de patiënten een stadium I of II ziekte. Een significant kleiner deel van de patiënten in de CWU+PET-groep onderging een onterechte thoractomie in vergelijking met de CWU-groep. De relatieve afname bedroeg 51% ($p = 0.003$) en de absolute afname 20% (41% van de thoracotomieen in de CWU-groep bleek onterecht). Dit resultaat betekent in de praktijk dat op elke 5 patiënten die een PET-scan ondergaan één onterechte thoracotomie wordt voorkomen. In 81% van de gevallen suggereerde PET terecht het doen van een thoracotomie en in 71% om er terecht van af te zien. Het belangrijkste effect van PET was het tonen van meer ziektelast, ofwel een hoger stadium.

Tegelijkertijd met de medische gegevens werden gegevens over kosten verzameld. Onder kosten werden gerekend, alle invasieve en niet-invasieve testen (inclusief PET), longoperaties en het aantal dagen ziekenhuisverblijf. De kosten werden berekend op individuele basis en op ziekenhuisniveau. De kosten met betrekking tot PET omvatte personele kosten, afschrijving- en onderhoudskosten, materiaal (FDG) en overhead. De kostprijs van een PET-scan varieerde van € 736 and € 1.588, afhankelijk van het type ziekenhuis en de FDG-beschikbaarheid (**hoofdstuk 6**).[9] In de 'CWU' groep bedroegen de gemiddelde kosten per patiënt € 9.573 en in de CWU+PET groep € 8.284. De belangrijkste kostenpost bleek het aantal hersteldagen in het ziekenhuis na chirurgie. Uit sensitiviteitsanalyses, op basis van variaties in de effectiviteit van PET (betrouwbaarheidsintervallen uit de PLUS studie), bleek dat de resultaten van de kostenanalyse robuust zijn in het voordeel van PET. In aanvulling op de kostenanalyse van de PLUS-studie onderzochten we drie scenario's waarbij verschillende parameters van PET en de ziekenhuissituatie werden gevarieerd. In een 'dure' variant werd uitgegaan van PET in een universitaire setting, met FDG-productie ter plekke en PET-scans voor zowel klinische- als onderzoeksdoeleinden. Als een 'goedkope' variant werd de situatie gekozen van een algemeen ziekenhuis waar 12 klinische PET-scans per dag kunnen worden gemaakt en de FDG van elders

wordt betrokken. De 'tussen-in' variant vertegenwoordigde een groot regionaal ziekenhuis waar FDG ter plekke beschikbaar is, 8 PET-scans per dag worden gemaakt en slechts beperkt onderzoek wordt verricht. Alle scenario's, uitgaande van de puntschatting van de effectiviteit van PET, vielen gunstig uit voor PET. Alleen bij de minste effectiviteit van PET (1 onterechte operatie voorkomen tegen 14 PET-scans) in de duurste variant (universiteit) bleek de conventionele aanpak goedkoper (namelijk € 542).

Niet alle experts onderschrijven de noodzaak voor randomisatie bij diagnostische studies.[10] Onnodig gebruik van middelen en ethische bezwaren worden zelfs gebruikt als argumenten tegen het doen van RCTs. Dit zette ons ertoe aan om onze ideeën over het nut en de noodzaak van gerandomiseerd diagnostisch onderzoek uiteen te zetten (**hoofdstuk 7**).[11] Samenvattend stellen wij dat de mate waarin een patiënt uiteindelijk voordeel heeft bij het toepassen van een diagnostisch instrument (in termen van iatrogene toxiciteit of verbetering in overleving) het best kan worden onderzocht door gelijktijdige vergelijking met de bestaande klinische praktijk zonder dat daar onrealistische of niet-verifieerbare aannames voor hoeven te worden gemaakt. De meest efficiënte strategie daartoe is, om de patiënten willekeurig te verdelen over de conventionele strategie met of zonder de nieuwe diagnostiek, waardoor bekende en deels onbekende maar bepalende factoren evenwichtig worden verdeeld en de test als enige variabele overblijft. Daarnaast is het van belang om voldoende patiënten in de studie op te nemen. Het opzetten van gerandomiseerde studies brengt bovendien andere voordelen met zich mee die niet zozeer met het randomiseren zelf te maken hebben, maar met intrinsieke factoren als transparantie, doordat een duidelijk protocol aan de studie ten grondslag ligt en het opstellen van hypotheses met bijbehorende berekeningen voor de steekproefgrootte.
De twee gerandomiseerde studies met FDG-PET bij longkanker laten zien dat dergelijke pragmatische RCTs uitvoerbaar zijn. In de PLUS studie werden 188 patiënten, met stadium I-III, in één jaar geincludeerd. De POORT studie, waarbij de rol van FDG-PET vroeg in het diagnostische traject werd gepositioneerd ter mogelijke vervanging van andere testen, werden in minder dan twee jaar tijd 465 patiënten aangemeld.[12] Het feit dat FDG-PET in die tijd nog weinig beschikbaar was heeft mogelijk wel bijgedragen aan dit succes.
Hoofdstuk 8 is een omvattend geheel van onze ervaring ten aanzien van de evaluatie van de kosteneffectiviteit van FDG-PET voor verschillende indicaties. Het vormt een raamwerk, wat in grote lijnen de hiërarchische aanpak voor de waardebepaling van diagnostische studies zoals eerder hierboven beschreven is, volgt. Het startpunt van onze studies was de geconstateerde inefficiëntie bij conventionele diagnostiek waarbij de betere accuratesse van FDG-PET, zoals gesuggereerd, mogelijk winst zou kunnen opleveren. Het retrospectieve cohortonderzoek toonde de aard van de problematiek en betekende een belangrijke motivering voor de clinici om aan verder onderzoek mee te werken. De 'aanloop'-studie naar de klinische waarde van PET vormde de leercurve voor diagnostici en behandelaars. Uiteindelijk droeg de opeenvolging en samenhang van de studies bij aan een vruchtbare samenwerking wat naar onze mening de RCTs en het ontwikkelen en implementeren van richtlijnen in de regio ten goede kwam.

Het heden

Anno 2006 is de PET-situatie in Nederland wezenlijk anders. De PET-scanner-capaciteit is enorm toegenomen en momenteel hebben de meeste ziekenhuizen toegang tot of de beschikking over een vaste of mobiele PET. In de Amsterdamse regio zijn in een straal van 20 kilometer tenminste 5 scanners. Vanaf 2004 bevatten de nationale richtlijnen voor diagnose en behandeling van niet-kleincellige longtumoren regels voor het gebruik van FDG-PET.[13] Voortdurend worden deze richtlijnen door middel van studies aangescherpt. Geschat wordt dat in ongeveer 80% van de patiënten, met een verdenking op operabele NSCLC, een PET-scan deel uit maakt van de reguliere diagnostiek. Een recente analyse van de gegevens uit de kankerregistratie van het Integraal Kankercentrum Amsterdam suggereert een absolute afname van 20% in het aantal longoperaties sinds 2000 (wat overeenkomt met een relatieve afname van 50% in onnodige operaties) vergeleken met de 5 voorgaande jaren.[14]

Sinds de opkomst van de PET-scanner hebben verbeteringen in het concept van de scanner en andere nieuwe technologieën het licht gezien. Na 15 jaar 'whole-body' PET is recentelijk een geïntegreerde of hybride PET-CT scanner geïntroduceerd.[15-17] Verschillende studies hebben superieure accuratesse gesuggereerd voor deze geïntegreerde PET-CT in vergelijking met het afzonderlijk visueel beoordelen van CT en PET-beelden (zoals gebruikelijk bij de PLUS-studie). Een eenvoudige zoekactie in MEDLINE leverde 5 studies op die de accuratesse van de geïntegreerde PET-CT scanner onderzochten in vergelijking met andere technieken in longkanker.[18-22] Alle 5 suggereerde een verbeterde accuratesse met PET-CT vergeleken met het afzonderlijk beoordelen van de beelden en twee studies claimden veranderingen in behandeling van patiënten op basis van de hybride PET-CT.[18,19] Echter, de eerste vraag die zou moeten worden gesteld is wat de aard en de mate van inefficiëntie is bij de huidige diagnostiek, waarbij inmiddels de PET wordt gebruikt. Hoewel we er vertrouwen in hebben dat het aantal longoperaties is afgenomen door de toevoeging van PET, weten we uit de PLUS studie dat nog steeds 20% van de operaties niet zinvol is. Om in te schatten of de geïntegreerde PET-CT iets van deze 20% onnodige operaties af had kunnen halen hebben we naar de individuele gevallen uit de PLUS studie gekeken: in 9 patiënten bleek de longoperatie zelf niet zinvol te zijn en in 10 andere patiënten kwam de ziekte vroeg terug of stierven de patiënten binnen één jaar. Van de 9 patiënten bij wie de operatie zelf al onnodig bleek hadden 2 een goedaardige aandoening, 6 hadden N2 ziekte of erger en 1 patiënt kon niet afdoende worden geopereerd wegens onvoldoende longcapaciteit. Van de 6 patiënten bij wie peroperatief een hoger een stadium werd aangetroffen bleek in 3 gevallen een mediastinoscopie de verdenking in de lymfeklieren, zoals gesuggereerd door de PET, niet te kunnen bevestigen. In de 10 patiënten waarbij in de follow-up duidelijk werd dat de chirurgie onnodig was, kwam bij 4 patiënten de ziekte vroeg terug na ogenschijnlijk curatieve chirurgie. Twee patiënten kregen botmetastasen en één hersen- en botmetastasen na het weigeren van een PET en bij de laatste patiënt bleken longmetastasen van een melanoom (primaire locatie onbekend) te zijn verwijderd waarna de ziekte weer terugkwam. Vijf van de 10 patiënten

overleden aan oorzaken gerelateerd aan de chirurgie en bij één was de doodsoorzaak onbekend. Concluderend had een betere niet-invasieve diagnostiek met PET-CT dus hooguit bij 3 van de 19 patiënten een onnodige operatie kunnen voorkomen. We realiseren ons dat deze aantallen te klein zijn voor definitieve conclusies, maar het geeft een indruk van de te verwachten winst van PET-CT bij het voorkomen van onnodige operaties.

Vergeleken met PET alleen, draagt PET-CT zeker bij aan de specificiteit van PET-interpretatie met als mogelijk klinisch gevolg dat biopsieën ter bevestiging van kwaadaardigheid efficiënter kunnen worden uitgevoerd. In de PLUS-studie bleken PET-positieve lagere cervicale lymfeklieren moeilijk te bevestigen door radiologen in de algemene ziekenhuizen. Echter, het mislukken van de bevestigende procedures heeft het aantal onnodige operaties niet beïnvloed, aangezien dergelijke lymfekliervertenking ook op andere manieren bevestigd kan worden (zij het wat invasiever). Wat betreft de waarde van PET-CT op de sensitiviteit: we hebben beredeneerd dat test-resultaten zelden tweeledig zijn en daarom kan PET-CT, ondanks dat het is opgebouwd uit dezelfde PET en CT-componenten als in de enkelvoudige opzet, het kwartje doen kantelen naar een hoger niveau van waarschijnlijkheid. Echter, wanneer de redenering ten aanzien van problemen die blijven na gebruik van PET op basis van de PLUS-studie klopt, dan verwachten we geen enorme verbetering hierin. Beweringen over verbeterde stadiering op het gebied van de primaire tumoruitbreiding[22] moeten nog worden bevestigd, gezien de beperkte resolutie van PET.[23] Alternatieve manieren om het mediastinum preoperatief te onderzoeken geven meer hoop voor de toekomst. Endoscopische echografie met naald punctie vanuit de slokdarm of vanuit de bronchus (EUS-FNA en EBUS-FNA) zijn nieuwe minimaal invasieve technieken met hoge potentie voor het beoordelen van klieren in het mediastinum en centraal gelegen tumoren.[24]

In het geval van PET-CT kunnen ook logistieke factoren de overstap naar deze modaliteit rechtvaardigen, zelfs wanneer verbetering op klinisch gebied of lagere kosten onwaarschijnlijk zijn, of moeilijk aan te tonen. Het concept van stadieren met één enkele scan ('one-stop-shop') in plaats van door een hele batterij aan testen over een langere periode is erg aantrekkelijk voor zowel de patiënt als vanuit het perspectief van de behandeling. Bovendien zal een kortere PET-scan tijd ook worden gewaardeerd door hen die momenteel met wachtlijsten te maken hebben.

Hoe het ook zij, welke methodologische bezwaren we ook mogen hebben ten aanzien van de bewijsvoering van de superioriteit van PET-CT ten opzicht van PET, de industriële ontwikkeling is nu eenmaal zo dat binnenkort de PET-CT de standaard is, eenvoudigweg omdat enkelvoudige PET-scanners niet meer worden verkocht.

Eén van de vragen die dan speelt, is of PET-CT aan het begin van diagnostisch traject moet worden ingezet of juist net voor het (invasieve) onderzoek van het mediastinum: het is niet economisch om een PET-CT-scan bij elke patiënt te maken, wanneer alleen al op basis van CT-bevindingen een groot aantal van de patiënten kunnen worden uitgesloten voor chirurgie. Modelmatige analyses, bij voorkeur gebaseerd op informatie over werkelijke kosten en scenarioanalyses kunnen bijdragen aan de oplossing van dit probleem.

Tot nu toe is FDG-PET aan bod gekomen in de context van chirurgische beslissingen enkel gebaseerd op TNM-stadiërings kwesties. Echter, de door PET verkregen metabole informatie lijkt ook prognostische informatie toe te voegen binnen de klinische stadia[25]: hoge opname door tumoren blijkt prognostisch ongunstig in vergelijking met lage opname. Ondanks dat het ontbreken van standaardisering van kwantitatieve PET-procedures, hetgeen het uitvoeren van meta-analyses van individuele studies bemoeilijkt, lijkt dit toch een consistente bevinding. Hoe deze informatie kan worden gecombineerd met andere prognostische factoren teneinde strategieën te ontwikkelen om de relatief slechte uitkomsten van patiënten met operabele NSCLC te verbeteren valt nog te bezien. Het komt waarschijnlijk neer op systemische therapieën voor specifieke subgroepen. Een vergelijkbaar prognostische aspect van PET is mogelijk relevant bij patiënten met lokaal uitgebreide ziekte die worden behandeld met combinatie therapieën. De rol van chirurgie staat hierbij ter discussie en mogelijk kunnen subgroepen van patiënten worden geïdentificeerd die baat hebben bij chirurgie. In observationele studies hebben wij, en anderen ook, laten zien dat het metabole gedrag, in beeld gebracht met PET, al vroeg tijdens systemische therapie, een voorspellende waarde kan hebben.[26,27] Om de klinische relevantie van dergelijke metabole patronen te bepalen zijn wederom gerandomiseerde onderzoeken nodig. Bij die onderzoeken verschuift de interesse van het bepalen van het effect van een diagnostische test op TNM-van gerelateerde uitkomsten, zoals onnodige chirurgie, naar die van het effect van diagnose-behandel-combinaties op de verbetering van overleving.

Met de volgende zinnen kan het proefschrift worden samengevat: voor het evalueren van (kosten-) effectiviteit van diagnostische technieken zijn studies nodig die zich richten op klinisch relevante eindpunten voorbij de accuraatheids parameters. Beargumenteerd en getoond wordt dat gerandomiseerde klinische studies in het bijzonder resultaten kunnen voortbrengen die overtuigen en direct toepasbaar zijn in de praktijk. Het raamwerk van studies dat hier wordt gepresenteerd kan van nut zijn bij het evalueren van andere diagnostische gereedschappen.

Referenties

1. Herder, GJ, Verboom, P., Smit, E. F. et al. Practice, efficacy and cost of staging suspected non-small cell lung cancer: a retrospective study in two Dutch hospitals. Thorax, 2002; 57: 11-14.
2. Flynn, K and Adams, E. Technology Assessment: Positron Emission Tomography. Original Report, 1996. 1996;
3. Freedman, LS. Evaluating and comparing imaging techniques: a review and classification of study designs. Br J Radiol, 1987; 60: 1071-1081.
4. Fryback, DG and Thornbury, J. R. The efficacy of diagnostic imaging. Med Decis Making, 1991; 11: 88-94.
5. Verboom, P, Herder, G. J., Hoekstra, O. S. et al. Staging of non-small-cell lung cancer and application of FDG-PET. A cost modeling approach. Int J Technol Assess Health Care, 2002; 18: 576-585.
6. Herder, GJ, van Tinteren, H., Comans, E. F. et al. Prospective use of serial questionnaires to evaluate the therapeutic efficacy of 18F-fluorodeoxyglucose (FDG) positron emission tomography (PET) in suspected lung cancer. Thorax, 2003; 58: 47-51.
7. van Tinteren, H, Hoekstra, O. S., Smit, E. F. et al. Toward less futile surgery in non-small cell lung cancer? A randomized clinical trial to evaluate the cost-effectiveness of positron emission tomography. Control Clin Trials, 2001; 22: 89-98.
8. van Tinteren, H, Hoekstra, O. S., Smit, E. F. et al. Effectiveness of positron emission tomography in the preoperative assessment of patients with suspected non-small-cell lung cancer: the PLUS multicentre randomised trial. Lancet, 2002; 359: 1388-1393.
9. Verboom, P, van Tinteren, H., Hoekstra, O. S. et al. Cost-effectiveness of FDG-PET in staging non-small cell lung cancer: the PLUS study. Eur J Nucl Med Mol Imaging, 2003; 30: 1444-1449.
10. Valk, PE. Do we need randomised trials to evaluate diagnostic procedures? Against. Eur J Nucl Med Mol Imaging, 2004; 31: 132-135.
11. van Tinteren, H, Hoekstra, O. S., and Boers, M. Do we need randomised trials to evaluate diagnostic procedures? For. Eur J Nucl Med Mol Imaging, 2004; 31: 129-131.
12. Herder, GJ, Kramer, H., Hoekstra, O. S. et al. Traditional Versus Up-Front [18F] Fluorodeoxyglucose-Positron Emission Tomography Staging of Non-Small-Cell Lung Cancer: A Dutch Cooperative Randomized Study. J Clin Oncol, 2006; 24: 1800-1806.
13. van Meerbeeck, JP, Koning, C. C., Tjan-Heijnen, V. C. et al. [Guideline on 'non-small cell lung carcinoma; staging and treatment']. Ned Tijdschr Geneeskd, 2005; 149: 72-77.
14. van Tinteren, H, Smit, E. F., and Hoekstra, O. S. FDG-PET in addition to conventional work-up in non-small-cell lung cancer. J Clin Oncol, 2005; 23: 1591-1592.
15. Beyer, T, Townsend, D. W., Brun, T. et al. A combined PET/CT scanner for clinical oncology. J Nucl Med, 2000; 41: 1369-1379.
16. Vogel, WV, Oyen, W. J., Barentsz, J. O. et al. PET/CT: panacea, redundancy, or something in between? J Nucl Med, 2004; 45 Suppl 1:15S-24S.: 15S-24S.
17. Townsend, DW, Carney, J. P., Yap, J. T. et al. PET/CT today and tomorrow. J Nucl Med, 2004; 45 Suppl 1:4S-14S.: 4S-14S.
18. Halpern, BS, Schiepers, C., Weber, W. A. et al. Presurgical staging of non-small cell lung cancer: positron emission tomography, integrated positron emission tomography/CT, and software image fusion. Chest, 2005; 128: 2289-2297.
19. Cerfolio, RJ, Ojha, B., Bryant, A. S. et al. The accuracy of integrated PET-CT compared with dedicated pet alone for the staging of patients with nonsmall cell lung cancer. Ann Thorac Surg, 2004; 78: 1017-1023.
20. Lardinois, D, Weder, W., Hany, T. F. et al. Staging of non-small-cell lung cancer with integrated positron-emission tomography and computed tomography. N Engl J Med, 2003; 348: 2500-2507.
21. Antoch, G, Stattaus, J., Nemat, A. T. et al. Non-small cell lung cancer: dual-modality PET/CT in preoperative staging. Radiology, 2003; 229: 526-533.
22. Keidar, Z, Haim, N., Guralnik, L. et al. PET/CT using 18F-FDG in suspected lung cancer recurrence: diagnostic value and impact on patient management. J Nucl Med, 2004; 45: 1640-1646.

23. Allen-Auerbach, M, Yeom, K., Park, J. et al. Standard PET/CT of the Chest During Shallow Breathing Is Inadequate for Comprehensive Staging of Lung Cancer. J Nucl Med, 2006; 47: 298-301.
24. Annema, JT, Hoekstra, O. S., Smit, E. F. et al. Towards a minimally invasive staging strategy in NSCLC: analysis of PET positive mediastinal lesions by EUS-FNA. Lung Cancer, 2004; 44: 53-60.
25. Vansteenkiste, J, Fischer, B. M., Dooms, C. et al. Positron-emission tomography in prognostic and therapeutic assessment of lung cancer: systematic review. Lancet Oncol, 2004; 5: 531-540.
26. Hoekstra, CJ, Stroobants, S. G., Smit, E. F. et al. Prognostic relevance of response evaluation using [18F]-2-fluoro-2-deoxy-D-glucose positron emission tomography in patients with locally advanced non-small-cell lung cancer. J Clin Oncol, 2005; 23: 8362-8370.
27. Cerfolio, RJ, Bryant, A. S., Winokur, T. S. et al. Repeat FDG-PET after neoadjuvant therapy is a predictor of pathologic response in patients with non-small cell lung cancer. Ann Thorac Surg, 2004; 78: 1903-1909.

List of publications

References from this thesis and associated papers in which the author was involved

Herder, G.J., Kramer, H., Hoekstra, O.S., Smit, E.F., Pruim, J., **van Tinteren, H.**, Comans, E.F., Verboom, P., Uyl-de Groot, C.A., Welling, A., Paul, M.A., Boers, M., Postmus, P.E., Teule, G.J., and Groen, H.J. Traditional versus up-front [18F] fluorodeoxyglucose-positron emission tomography staging of non-small-cell lung cancer: a Dutch cooperative randomized study. J Clin Oncol. 2006; 12: 1800-1806.

van Tinteren, H., Hoekstra, O.S., Smit, E.F., and Boers, M. The implementation of PET in non-small-cell lung cancer in the Netherlands. Clin Oncol (R Coll Radiol). 2006; 2: 156-157.

Hoekstra, C.J., Stroobants, S.G., Smit, E.F., Vansteenkiste, J., **van Tinteren, H.**, Postmus, P.E., Golding, R.P., Biesma, B., Schramel, F.J., van Zandwijk, N., Lammertsma, A.A., and Hoekstra, O.S. Prognostic relevance of response evaluation using [18F]-2-fluoro-2-deoxy-D-glucose positron emission tomography in patients with locally advanced non-small-cell lung cancer. J Clin Oncol. 2005; 33: 8362-8370.

Herder, G.J., **van Tinteren, H.**, Golding, R.P., Kostense, P.J., Comans, E.F., Smit, E.F., and Hoekstra, O.S. Clinical prediction model to characterize pulmonary nodules: validation and added value of 18F-fluorodeoxyglucose positron emission tomography. Chest. 2005; 4: 2490-2496.

van Tinteren, H., Smit, E.F., and Hoekstra, O.S. FDG-PET in addition to conventional work-up in non-small-cell lung cancer. J Clin Oncol. 2005; 7: 1591-1592.

van Tinteren, H., Hoekstra, O.S., and Boers, M. Do we need randomised trials to evaluate diagnostic procedures? For. Eur J Nucl Med Mol Imaging. 2004; 1: 129-131.

Verboom, P., **van Tinteren, H.**, Hoekstra, O.S., Smit, E.F., van den Bergh, J.H., Schreurs, A.J., Stallaert, R.A., van Velthoven, P.C., Comans, E.F., Diepenhorst, F.W., van Mourik, J.C., Postmus, P.E., Boers, M., Grijseels, E.W., Teule, G.J., and Uyl-de Groot, C.A. Cost-effectiveness of FDG-PET in staging non-small cell lung cancer: the PLUS study. Eur J Nucl Med Mol Imaging. 2003; 11: 1444-1449.

van Tinteren, H., Hoekstra, O.S., and Boers, M. The need for Health Technology Assessments of PET. Eur J Nucl Med Mol Imaging. 2003; 10: 1438-1439.

Hoekstra, C.J., Stroobants, S.G., Hoekstra, O.S., Vansteenkiste, J., Biesma, B., Schramel, F.J., van Zandwijk, N., **van Tinteren, H.**, and Smit, E.F. The value of [18F]fluoro-2-deoxy-D-glucose positron emission tomography in the selection of patients with stage IIIA-N2 non-small cell lung cancer for combined modality treatment. Lung Cancer. 2003; 2: 151-157.

Herder, G.J., **van Tinteren, H.**, Comans, E.F., Hoekstra, O.S., Teule, G.J., Postmus, P.E., Joshi, U., and Smit, E.F. Prospective use of serial questionnaires to evaluate the therapeutic efficacy of 18F-fluorodeoxyglucose (FDG) positron emission tomography (PET) in suspected lung cancer. Thorax. 2003; 1: 47-51.

van Tinteren, H., Hoekstra, O.S., Smit, E.F., van den Bergh, J.H., Schreurs, A.J., Stallaert, R.A., van Velthoven, P.C., Comans, E.F., Diepenhorst, F.W., Verboom, P., van Mourik, J.C., Postmus, P.E., Boers, M., and Teule, G.J. Effectiveness of positron emission tomography in the preoperative assessment of patients with suspected non-small-cell lung cancer: the PLUS multicentre randomised trial. Lancet. 2002; 9315: 1388-1393.

van Tinteren, H., Hoekstra, O.S., Smit, E.F., Verboom, P., and Boers, M. Toward less futile surgery in non-small cell lung cancer? A randomized clinical trial to evaluate the cost-effectiveness of positron emission tomography. Control Clin Trials. 2001; 1: 89-98.

Dankwoord

Dankwoord

Dit multidisciplinaire project is dankzij de inhoudelijke en ondersteunende hulp van velen tot stand gekomen. Zonder daarbij iemand tekort te willen doen, zou ik een paar mensen in het bijzonder willen bedanken. Allereerst het (voormalige) duo Professor Jaap Teule en Professor Otto Hoekstra. Samen hebben ze op visionaire wijze het klinische gebruik van de PET internationale bekendheid gegeven hetgeen op lokaal niveau inmiddels merkbaar is. Otto, een FDG-plaatje van jou kan alleen maar één groot zwart beeld opleveren. Professor Maarten Boers, hoewel je kwam nadat een deel van de PET-onderzoeken van start waren gegaan ben je van meet af aan een enthousiaste pleitbezorger van de ontwerpen geweest. Je kritische blik bracht de stukken op een hoger plan. Professor Egbert Smit heeft vanuit de discipline longziekten zonder veel woorden een essentiële rol gespeeld. Professor Carin Uyl en Paul Verboom brachten met miraculeuze spreadsheets de kosten van PET in beeld. Judith Herder was alle jaren klankbord en deelde haar ervaring op het gebied van promoveren. Hoewel ik 'van buiten' kwam heb ik veel collegialiteit ondervonden van medewerkers van de afdeling nucleaire geneeskunde. Emile Comans, Arthur van Lingen, professor Adriaan Lammertsma en het secretariaat met Amanda en Jaap om er een paar te noemen. Dank ben ik ook verschuldigd aan alle artsen en patiënten die hebben bijgedragen aan de studies. Dankzij de brede samenwerking werd het een groot succes en heeft in het bijzonder de PLUS-studie internationale faam verworven. De in het begin wat onwennige interdisciplinaire onderzoekers-bijeenkomsten liepen vooral onder voorzitterschap van Piet van Veldhoven als 'eminent grise' steeds goed af. De rijsttafels deden de rest. Willeke Heibroek en Tinie Benraadt hebben een bijzonder faciliterende rol gespeeld. Hoewel veel van het werk niet direct tot de 'core-business' van het IKA behoorde kreeg ik van jullie de ruimte en het vertrouwen. Fred Diepenhorst en vele andere collega's van het IKA hebben zich altijd enthousiast en deskundig ingezet bij de dataverzameling van de PLUS-, en later de POORT-studie. Otilia Dalesio beschouw ik als mijn algemene mentor. Mijn kennis, manier van werken en houding ten aanzien van klinisch onderzoek zijn grotendeels door haar gevormd. Tenslotte dank ik mijn vrienden en familie en in het bijzonder mijn ouders, die waarschijnlijk nauwelijks doorhadden waarmee ik bezig was, maar altijd vertrouwen in mij hebben getoond.

Curriculum vitae

Curriculum vitae

Harm van Tinteren was born in 1962 in Oranjestad (Aruba, NA). He studied at the Agricultural University (now University of Life Sciences) in Wageningen, the Netherlands. He graduated in 1988 on majors in Human Nutrition Epidemiology, Statistics and Informatics (cum laude). Since 1988 he works as a statistician at the Biometrics Department of the Netherlands Cancer Institute and from 1990 he also fulfills a part-time position at the Comprehensive Cancer Center in Amsterdam, the Netherlands. He is (co-)author of about 100 publications and focuses on the design and processing of clinical trials in cancer and diagnostics. He is a registered biostatistician of the Netherlands Society for Statistics and Operations Research (VVS).